

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**



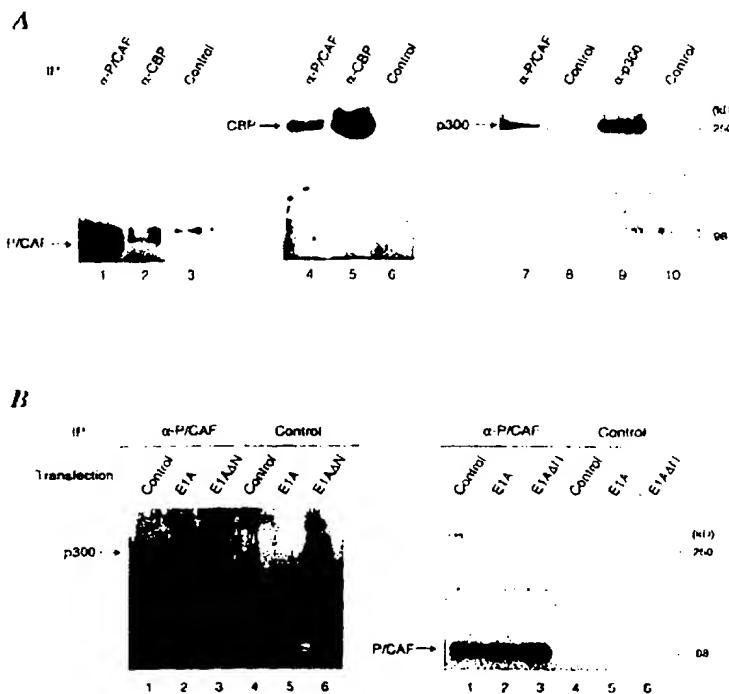
## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> : <b>C12N 15/12, C07K 14/47, G01N 33/50, A61K 38/17</b>		<b>A2</b>	(11) International Publication Number: <b>WO 98/03652</b>
			(43) International Publication Date: 29 January 1998 (29.01.98)
(21) International Application Number: <b>PCT/US97/12877</b>		(74) Agents: MILLER, Mary, L. et al.; Needle & Rosenberg, Suite 1200, 127 Peachtree Street, N.E., Atlanta, GA 30303 (US).	
(22) International Filing Date: 23 July 1997 (23.07.97)			
(30) Priority Data: 60/022,273      23 July 1996 (23.07.96)      US		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).	
(60) Parent Application or Grant (63) Related by Continuation US      60/022,273 (CIP) Filed on      23 July 1996 (23.07.96)			
(71) Applicant (for all designated States except US): THE GOV- ERNMENT OF THE UNITED STATES OF AMERICA, represented by THE SECRETARY, DEPARTMENT OF HEALTH AND HUMAN SERVICES [US/US]; National Institutes of Health, Office of Technology Transfer Suite 325, 6011 Executive Boulevard, Rockville, MD 20852 (US).		Published Without international search report and to be republished upon receipt of that report.	
(72) Inventors; and (75) Inventors/Applicants (for US only): NAKATANI, Yoshihiro [JP/US]; 4624 Edgefield Road, Bethesda, MD 20814 (US). HOWARD, Bruce, H. [US/US]; 8715 Fallen Oak Drive, Bethesda, MD 20817 (US).			

(54) Title: P300/CBP-ASSOCIATED TRANSCRIPTIONAL CO-FACTOR P/CAF AND USES THEREOF

## (57) Abstract

The present invention provides a purified protein designated P/CAF having a molecular weight of about 93,000 daltons as determined by sodium dodecyl sulfate polyacrylamide gel electrophoresis under reducing conditions and which acetylates histones and which also binds to the p300/CBP cellular protein. The present invention further provides a nucleic acid encoding the P/CAF protein as well as a vector containing the nucleic acid and a host for the vector. A purified antibody which specifically binds the P/CAF protein is also provided. Also provided are methods of screening for compounds that inhibit or stimulate the transcription modulating and histone acetyltransferase activity of P/CAF and p300/CBP.



**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

## P300/CBP-ASSOCIATED TRANSCRIPTIONAL CO-FACTOR P/CAF AND USES THEREOF

**BACKGROUND OF THE INVENTION**

5

**Field of the Invention**

The present invention provides a transcriptional co-factor, p300/CBP-associated factor (P/CAF), which modulates transcription through binding to the cellular transcription co-factors p300 and CBP and through acetylation of histones. Also  
10 provided are methods for screening for the presence of P/CAF and for substances which alter the transcription modulating effect and growth regulatory activity of P/CAF.

**Background Art**

Cellular proteins p300 and CBP are global transcriptional coactivators that are  
15 involved in the regulation of various DNA-binding transcriptional factors (Janknecht and Hunter, 1996). Recently, p300 was found to be very closely related to CBP, a factor that binds selectively to the protein kinase A-phosphorylated form of CREB (3-5). Cellular factors p300 and CBP exhibit strong amino acid sequence similarity and share the capacity to bind both CREB and E1A (6-8). Although neither p300 nor CBP by  
20 itself binds to DNA, each can be recruited to promoter elements via interaction with sequence-specific activators and functions to be a transcriptional adaptor. For simplicity, p300 and CBP will be termed p300/CBP in the context of discussing their shared functional properties.

25 p300/CBP is a large protein consisting of over 2,400 amino acids, known to interact with a variety of DNA-binding transcriptional factors including nuclear hormone receptors (13,57), CREB (3,4, 7), c-Jun/v-Jun (9,11), YY1 (10), c-Myb/v-Myb (12,58), Sap-1a (59), c-Fos (11) and MyoD (60). DNA-binding factors recruit p300/CBP not only by direct but also indirect interactions through cofactors; for example, nuclear  
30 hormone receptors recruit p300/CBP directly as well as through indirect interactions, via SRC-1, which stimulates transcription by binding to various nuclear hormone receptors (13,61).

The transforming proteins encoded by adenovirus and several other small DNA tumor viruses disturb host cell growth control by interacting with cellular factors that normally function to repress cell proliferation. One of the most intensively studied of these viral proteins, the product of the adenovirus E1A gene, is itself sufficient for transformation (1). E1A transforming activity resides in two distinct domains, the targets of which include p300/CBP and products of the retinoblastoma (RB) susceptibility gene family (1,2). Interactions of E1A with p300/CBP and RB are thought to influence functionally distinct growth regulatory pathways, allowing the two domains to contribute additively to transformation (1).

10

The paradigm for how E1A and functionally related viral proteins perturb cell growth regulation derives in large part from studies on their interactions with RB (1,2). The molecular function of E1A is based on its capacity to interfere with cellular protein-protein interactions. Since both E1A and various cellular targets bind to a site in RB termed the pocket domain (2), E1A can competitively disrupt the complex formation between RB and its cellular targets.

15

The second cellular factor implicated in E1A-dependent transformation, p300, is believed to inhibit G0/G1 exit, to activate certain enhancers, and to stimulate differentiation (1,2). E1A inhibits the p300/CBP-mediated transcriptional activation of many promoters (14). In one case that has been examined, the complex of p300 and YY1, E1A inhibits transcription without disrupting the complex (10).

20

The present invention provides a cellular protein designated P/CAF which binds to p300/CBP and plays an important role in both transcription and cell cycle regulation associated with a histone acetyltransferase activity. The present invention also provides a histone acetyltransferase activity in the p300/CBP cellular protein, thus providing targets for modulating transcription and cell cycle regulation in cells.

25

30

## SUMMARY OF THE INVENTION

The present invention provides a purified protein designated P/CAF having a molecular weight of about 93,000 daltons as determined by sodium dodecyl sulfate polyacrylamide gel electrophoresis under reducing conditions and which acetylates histones and which also binds to the p300/CBP cellular protein.

The present invention further provides a nucleic acid encoding the P/CAF protein as well as a vector containing the nucleic acid and a host for the vector. A purified antibody which specifically binds the P/CAF protein is also provided.

In addition, also provided is a bioassay for screening substances for the ability to inhibit the transcription modulating activity of P/CAF and/or histone acetyltransferase activity, comprising contacting the substance with a system in which histone acetylation by P/CAF can be determined; determining the amount of histone acetylation by P/CAF in the presence of the substance; and comparing the amount of histone acetylation by P/CAF in the presence of the substance with the amount of histone acetylation by P/CAF in the absence of the substance, a decreased amount of histone acetylation by P/CAF in the presence of the substance indicating a substance that can inhibit the transcription modulating activity and/or histone acetyltransferase activity of P/CAF.

Furthermore, the present invention provides a bioassay for screening substances for the ability to inhibit the transcription modulating activity and/or histone acetyltransferase activity of P/CAF comprising contacting the substance with a system in which the p300 binding of P/CAF can be determined; determining the amount of p300 binding of P/CAF in the presence of the substance; and comparing the amount of p300 binding of P/CAF in the presence of the substance with the amount of p300 binding of P/CAF in the absence of the substance, a decreased amount of p300 binding of P/CAF in the presence of the substance indicating a substance that can inhibit the transcription modulating activity and/or histone acetyltransferase activity of P/CAF.

Also provided is a method for determining the amount of P/CAF in a biological sample comprising contacting the biological sample with a polypeptide comprising the amino acid sequence of SEQ ID NO:3 under conditions whereby a P/CAF/p300 complex can be formed; and determining the amount of the P/CAF/p300 complex; the  
5 amount of the complex indicating the amount of P/CAF in the sample.

The present invention additionally provides a method for determining the amount of P/CAF in a biological sample comprising contacting the biological sample with an antibody which specifically binds P/CAF under conditions whereby a P/CAF/antibody  
10 complex can be formed; and determining the amount of the P/CAF/antibody complex, the amount of the complex indicating the amount of P/CAF in the sample.

Also provided herein is an assay for screening substances for the ability to inhibit or stimulate the histone acetyltransferase activity of P/CAF, comprising: contacting the  
15 substance with a system in which histone acetylation by P/CAF can be determined; determining the amount of histone acetylation by P/CAF in the presence of the substance; and comparing the amount of histone acetylation by P/CAF in the presence of the substance with the amount of histone acetylation by P/CAF in the absence of the substance, a decreased or increased amount of histone acetylation by P/CAF in the  
20 presence of the substance indicating a substance that can inhibit or stimulate, respectively, the histone acetyltransferase activity of P/CAF.

The present invention further provides an assay for screening substances for the ability to inhibit binding of P/CAF to p300/CBP comprising: contacting the substance  
25 with a system in which the P/CAF binding of p300/CBP can be determined; determining the amount of P/CAF binding of p300/CBP in the presence of the substance; and comparing the amount of binding of P/CAF to p300/CBP in the presence of the substance with the amount of binding of P/CAF to p300/CBP in the absence of the substance, a decreased amount of binding of P/CAF to p300/CBP in the presence of the  
30 substance indicating a substance that can inhibit the ability to inhibit binding of P/CAF to p300/CBP.

In addition, an assay is provided for screening substances for the ability to inhibit or stimulate the histone acetyltransferase activity of p300/CBP, comprising: contacting the substance with a system in which histone acetylation by p300/CBP can be  
5 determined; determining the amount of histone acetylation by p300/CBP in the presence of the substance; and comparing the amount of histone acetylation by p300/CBP in the presence of the substance with the amount of histone acetylation by p300/CBP in the absence of the substance, a decreased or increased amount of histone acetylation by  
10 p300/CBP in the presence of the substance indicating a substance that can inhibit or stimulate, respectively, the histone acetyltransferase activity of p300/CBP.

Furthermore, the present invention provides an assay for screening substances for the ability to inhibit binding of a DNA-binding transcription factor to p300/CBP comprising: contacting the substance with a system in which the DNA-binding  
15 transcription factor binding of P300/CBP can be determined; determining the amount of DNA-binding transcription factor binding of p300/CBP in the presence of the substance; and comparing the amount of binding of DNA-binding transcription factor to p300/CBP in the presence of the substance with the amount of binding of DNA-binding  
20 transcription factor to p300/CBP in the absence of the substance, a decreased amount of binding of DNA-binding transcription factor to p300/CBP in the presence of the substance indicating a substance that can inhibit the ability to inhibit binding of DNA-binding transcription factor to p300/CBP.

A method is also provided for inhibiting the transcription modulating activity of  
25 P/CAF in a subject, comprising administering to the subject a transcription modulating activity inhibiting amount of a substance in a pharmaceutically acceptable carrier.

Also provided in the present invention is a method for stimulating the transcription modulating activity of P/CAF in a subject, comprising administering to the  
30 subject a transcription modulating activity stimulating amount of a substance in a pharmaceutically acceptable carrier.

Furthermore, the present invention provides a method for inhibiting the histone acetyltransferase activity of p300/CBP in a subject, comprising administering to the subject a histone acetyltransferase activity inhibiting amount of a substance in a  
5 pharmaceutically acceptable carrier.

Finally, the present invention additionally provides a method for stimulating the histone acetyltransferase activity of p300/CBP in a subject, comprising administering to the subject a histone acetyltransferase activity stimulating amount of a substance in a  
10 pharmaceutically acceptable carrier.

### BRIEF DESCRIPTION OF THE FIGURES

Figs. 1A-B. Fig 1A: P/CAF-p300/CBP interaction *in vivo*. Cell extract was  
15 immunoprecipitated with rabbit anti-P/CAF (lanes 1, 4, and 7), rabbit anti-CBP (lanes 2 and 5), and mouse anti-p300 (lane 9) antibodies. For controls, cell extract was precipitated with rabbit control IgG (lanes 3, 6, and 8) or mouse anti-HA monoclonal antibody (lane 10). The precipitates were analyzed by immunoblotting with anti-P/CAF (lanes 1-3), anti-CBP (lanes 4-6), and anti-p300 (lanes 7-10) antibodies. The positions  
20 of non-specific bands are indicated by asterisks. Fig. 1B: E1A inhibits the P/CAF-p300 interaction *in vivo*. Osteosarcoma cells were transfected with either control vector (lanes 1 and 4) or E1A- (lanes 2 and 5) or E1A $\Delta$ N- (lanes 3 and 6) expression vectors. Extract from the transfected subpopulation was immunoprecipitated with anti-P/CAF (lanes 1-3) or control (lanes 4-6) IgG. The precipitates were analyzed by  
25 immunoblotting with anti-p300 and anti-P/CAF.

Figs. 2A-F. P/CAF and E1A mediate antagonistic effects on cell cycle progression. HeLa cells (ATCC accession number CCL 2) were transfected by electroporation with 7  $\mu$ g of P/CAF-expression plasmid and/or 3  $\mu$ g of the full-length or  
30 the N-terminally deleted ( $\Delta$ 2-36) E1A 12S-expression plasmid as indicated in the figure. These plasmids were constructed by subcloning FLAG-P/CAF and E1A cDNAs into

pCX (34) and pcDNA1 (Invitrogen), respectively. All samples, in addition, contained 1  $\mu$ g of sorting plasmid (pCMV-IL2R) (31) and carrier plasmid (pCX) to normalize the total amount of DNA to 11  $\mu$ g. After transfection, cells were incubated in Dulbecco's modified Eagle's medium with 10% fetal bovine calf serum for 12 hours and subsequently labeled in medium containing 10  $\mu$ M bromo-deoxyuridine (BrdU) for 30 min. Subsequently, the transfected subpopulation was purified by magnetic affinity cell sorting and nuclei were analyzed by dual parameter flow cytometry as described (32). Histograms show percentages of cells in G1 and S phases. Abscissa values represent fluorescence intensity of bound anti-BrdU antibodies in log scale.

10

Fig. 3. Histone acetyltransferase activity of P/CAF. Activity of hGCN5 (lanes 1 and 4) and P/CAF (lanes 2 and 5) that acetylates free histones (lanes 1-3) or histones in the nucleosome core particle (35) (lanes 4-6) was measured as described (36). Each reaction contains 0.3 pmol of affinity purified FLAG-hGCN5 or FLAG-P/CAF, 4 pmol of the histone octamer or the nucleosome core particle and 10 pmol of [1- $^{14}$ C]acetyl-CoA. Note that the histone octamer dissociates into dimers or tetramers under assay conditions. Acetylated histones were detected by autoradiography after separation by SDS-PAGE. The bands corresponding to acetylated histones H3 and H4 are indicated by arrows.

20

## DETAILED DESCRIPTION OF THE INVENTION

As used in the specification and in the claims, "a" can mean one or more, depending upon the context in which it is used.

25

### P/CAF protein and fragments

The present invention provides a purified protein designated P/CAF having a molecular weight of about 93,000 daltons as determined by sodium dodecyl sulfate polyacrylamide gel electrophoresis under reducing conditions and which acetylates histones. The P/CAF protein can also bind to the amino acid region of SEQ ID NO:3

30

(amino acid (aa) residues 1753 - 1966) of the cellular transcriptional factor, p300 (which has the complete amino acid sequence of SEQ ID NO:6 and the nucleotide sequence of SEQ ID NO:12), and the amino acid region of SEQ ID NO:6 (amino acid residues 1805 - 1854) of the cellular transcriptional factor, CBP (which has the complete amino acid sequence of SEQ ID NO:7 and the nucleotide sequence of SEQ ID NO:13). The P/CAF protein can be defined by any one or more of the typically used parameters. Examples of these parameters include, but are not limited to molecular weight (calculated or empirically determined), isoelectric focusing point, specific epitope(s), complete amino acid sequence, sequence of a specific region (e.g., N-terminus) of the amino acid sequence and the like.

For example, The P/CAF protein can consist of the amino acid sequence of SEQ ID NO:1 or the P/CAF protein can comprise the amino acid sequence of SEQ ID NO:2 which represents the carboxy terminal end of the P/CAF protein and contains the histone acetyltransferase activity, or the amino acid sequence of SEQ ID NO:4, which represents the amino terminal end of the P/CAF protein, containing the binding site for p300/CBP. Because the amino-terminal region is specific for P/CAF it can be used to define and identify P/CAF.

As used herein, "purified" refers to a protein (polypeptide, peptide, etc.) that is sufficiently free of contaminants or cell components with which it normally occurs to distinguish it from the contaminants or other components of its natural environment. The purified protein need not be homogeneous, but must be sufficiently free of contaminants to be useful in a clinical or research setting, for example, in an assay for detecting antibodies to the protein. Greater levels of purity can be obtained using methods derived from well known protocols. Specific methods for purifying P/CAF proteins are known in the art.

As will be appreciated by those skilled in the art, the invention also includes those P/CAF polypeptides having slight variations in amino acid sequence which yield polypeptides equivalent to the P/CAF protein defined herein. Such variations may arise

naturally as allelic variations (*e.g.*, due to genetic polymorphism) or may be produced by human intervention (*e.g.*, by mutagenesis of cloned DNA sequences), such as induced point, deletion, insertion and substitution mutants. Minor changes in amino acid sequence are generally preferred, such as conservative amino acid replacements, small internal deletions or insertions, and additions or deletions at the ends of the molecules. Substitutions may be designed based on, for example, the model of Dayhoff, *et al.* (37). These modifications can result in changes in the amino acid sequence, provide silent mutations, modify a restriction site, or provide other specific mutations.

- 10           Modifications to any of the P/CAF proteins or fragments can be made, while preserving the specificity and activity (function) of the native protein or fragment thereof. As used herein, "native" describes a protein that occurs in nature. The modifications contemplated herein can be conservative amino acid substitutions, for example, the substitution of a basic amino acid for a different basic amino acid.
- 15           Modifications can also include creation of fusion proteins with epitope tags or known recombinant proteins or genes encoding them created by subcloning into commercial or non-commercial vectors (*e.g.*, polyhistidine tags, flag tags, myc tag, glutathione-S-transferase [GST] fusion protein, *xyIE* fusion reporter construct). Furthermore, the modifications can be such as do not affect the function of the protein or the way the
- 20           protein accomplishes that function (*e.g.*, its secondary structure or the ultimate result of the protein's activity). These products are equivalent to the P/CAF protein. The means for determining the function, way and result parameters are well known.

- Having provided an example of a purified P/CAF protein, the invention also
- 25           enables the purification of P/CAF homologs from other species and allelic variants from individuals within a species. For example, an antibody raised against the exemplary human P/CAF protein can be used routinely to screen preparations from different humans for allelic variants of the P/CAF protein that react with the P/CAF protein-specific antibody. Similarly, an antibody raised against an epitope, for example, from a
- 30           conserved amino acid region of the human P/CAF protein can be used to routinely screen for homologs of the P/CAF protein in other species. A P/CAF protein can be

routinely identified in and obtained from other species and from individuals within a species using the methods taught herein and others known in the art. For example, given the present sequence, the DNA encoding a conserved amino acid sequence can be used to probe genomic DNA or DNA libraries of an organism to predictably obtain the P/CAF gene for that organism. The gene can then be cloned and expressed as the P/CAF protein and purified according to any of a number of routine, predictable methods. An example of the routine protein purification methods available in the art can be found in Pei et al. (38).

10 A purified polypeptide fragment of the P/CAF protein is also provided. The term "fragment" as used herein regarding a P/CAF protein, means a molecule of at least five contiguous amino acids of P/CAF protein that has at least one function shared by P/CAF protein or a region thereof. These functions can include antigenicity, binding capacity, acetyltransferase activity and structural roles, among others. The P/CAF  
15 fragment can be specific for a recited source. As used herein to describe an amino acid sequence (protein, polypeptide, peptide, etc.), "specific" means that the amino acid sequence is not found identically in any other source. The determination of specificity is made routine by the availability of computerized amino acid sequence databases and sequence comparison programs, wherein an amino acid sequence of almost any length  
20 can be quickly and reliably checked for the existence of identical sequences. If an identical sequence is not found, the protein is "specific" for the recited source. For example, a P/CAF fragment can be species-specific (e.g., found in the P/CAF protein of humans, but not of other species).

25 A fragment of the P/CAF protein having histone acetyltransferase activity can consist of the amino acid sequence of SEQ ID NO:2. A fragment of the P/CAF protein which binds to the amino acid sequence of SEQ ID NO:3 on p300 and the amino acid sequence of SEQ ID NO:9 on CBP can consist of the amino acid sequence of SEQ ID NO:4. To the extent that these fragments are specific for P/CAF, they can be used to  
30 identify and define P/CAF.

An antigenic fragment of P/CAF protein is provided. An antigenic fragment has an amino acid sequence of at least about five consecutive amino acids of a P/CAF protein amino acid sequence and binds an antibody or elicits an immune response in an animal. An antigenic fragment can be selected by applying the routine technique of epitope mapping to P/CAF protein to determine the regions of the proteins that contain epitopes reactive with antibodies or are capable of eliciting an immune response in an animal. Once the epitope is selected, an antigenic polypeptide containing the epitope can be synthesized directly, or produced recombinantly by cloning nucleic acids encoding the antigenic polypeptide in an expression system, according to standard methods.

Alternatively, an antigenic fragment of the antigen can be isolated from the whole P/CAF protein or a larger fragment of the P/CAF protein by chemical or mechanical disruption. Fragments can also be randomly chosen from a known P/CAF protein sequence and synthesized. The purified fragments thus obtained can be tested to determine their antigenicity and specificity by routine methods.

#### **Nucleic Acids Encoding P/CAF Protein**

An isolated nucleic acid that encodes a P/CAF protein is also provided. As used herein, the term "isolated" means a nucleic acid separated or substantially free from at least some of the other components of the naturally occurring organism, for example, the cell structural components commonly found associated with nucleic acids in a cellular environment and/or other nucleic acids. The isolation of nucleic acids can therefore be accomplished by techniques such as cell lysis followed by phenol plus chloroform extraction, followed by ethanol precipitation of the nucleic acids (39). It is not contemplated that the isolated nucleic acids are necessarily totally free of all non-nucleic acid components or all other nucleic acids, but that the isolated nucleic acids are isolated to a degree of purification to be useful in clinical, diagnostic, experimental, or other procedures such as, for example, gel electrophoresis, Southern, Northern or dot blot hybridization, or polymerase chain reaction (PCR).

A skilled artisan in the field will readily appreciate that there are a multitude of procedures which may be used to isolate the nucleic acids prior to their use in other procedures. These include, but are not limited to, lysis of the cell followed by gel filtration or anion exchange chromatography, binding DNA to silica in the form of glass  
5 beads, filters or diatoms in the presence of high concentrations of chaotropic salts, or ethanol precipitation of the nucleic acids.

The nucleic acids of the present invention can include positive and negative strand RNA as well as DNA and can include genomic and subgenomic nucleic acids  
10 found in the naturally occurring organism. The nucleic acids contemplated by the present invention include double stranded and single stranded DNA of the genome, complementary positive stranded cRNA and mRNA, and complementary cDNA produced therefrom and any nucleic acid which can selectively or specifically hybridize to the isolated nucleic acids provided herein. Stringent conditions (further described  
15 below) are used to distinguish selectively or specifically hybridizing nucleic acids from non-selectively and non-specifically hybridizing nucleic acids.

An isolated nucleic acid that encodes a P/CAF protein can be species-specific (i.e., does not encode the P/CAF protein of other species and does not occur in other  
20 species). Examples of the nucleic acids contemplated herein include the nucleic acid of SEQ ID NO:10 as well as the nucleic acids that encode each of the P/CAF proteins or fragments thereof described herein. P/CAF proteins and protein fragments can be routinely obtained as described herein and their structure (sequence) determined by routine means including the methods as used herein.

25

P/CAF protein-encoding nucleic acids can be isolated from an organism in which they are normally found (e.g., humans), using any of the routine techniques. For example, a genomic DNA or cDNA library can be constructed and screened for the presence of the nucleic acid of interest using one of the present P/CAF protein-encoding  
30 nucleic acids as a probe. Methods of constructing and screening such libraries are well known in the art and kits for performing the construction and screening steps are

commercially available (for example, Stratagene Cloning Systems, La Jolla, CA). Once isolated, the nucleic acid can be directly cloned into an appropriate vector, or if necessary, be modified to facilitate the subsequent cloning steps. Such modification steps are routine, an example of which is the addition of oligonucleotide linkers, which  
5 contain restriction sites, to the termini of the nucleic acid (See, for example, ref. 39).

P/CAF protein-encoding nucleic acids can also be synthesized. For example, a method of obtaining a DNA molecule encoding a specific P/CAF protein is to synthesize a recombinant DNA molecule which encodes the P/CAF protein. For example, nucleic  
10 acid synthesis procedures are routine in the art and oligonucleotides coding for a particular protein region are readily obtainable through automated DNA synthesis. A nucleic acid for one strand of a double-stranded molecule can be synthesized and hybridized to its complementary strand. One can design these oligonucleotides such that the resulting double-stranded molecule has either internal restriction sites or appropriate  
15 5' or 3' overhangs at the termini for cloning into an appropriate vector.

Oligonucleotides complementary to or identical with the P/CAF protein-encoding nucleic acid sequence can be synthesized as primers for amplification reactions, such as PCR, or as probes to detect P/CAF protein encoding nucleic acids by  
20 various hybridization protocols (e.g., Northern blot, Southern blot, dot blot, colony screening, etc.).

Double-stranded molecules coding for relatively large proteins can readily be synthesized by first constructing several different double-stranded molecules that code  
25 for particular regions of the protein, followed by ligating these DNA molecules together. For example, Cunningham, et al. (40), have constructed a synthetic gene encoding the human growth hormone by first constructing overlapping and complementary synthetic oligonucleotides and ligating these fragments together. See also, Ferretti, et al. (41), wherein synthesis of a 1057 base pair synthetic bovine rhodopsin gene from synthetic  
30 oligonucleotides is disclosed.

By constructing a P/CAF protein-encoding nucleic acid in this manner, one skilled in the art can readily obtain any particular P/CAF protein with modifications at any particular position or positions. See also, U.S. Patent No. 5,503,995 which describes an enzyme template reaction method of making synthetic genes. Techniques  
5 such as this are routine in the art and are well documented. DNA encoding the P/CAF protein or P/CAF protein fragments can then be expressed *in vivo* or *in vitro*.

The nucleic acid encoding the P/CAF protein can be any nucleic acid that functionally encodes the P/CAF protein. To functionally encode the protein (i.e., allow  
10 the nucleic acid to be expressed), the nucleic acid can include, but is not limited to, expression control sequences, such as an origin of replication, a promoter, regions upstream or downstream of the promoter, such as enhancers that may regulate the transcriptional activity of the promoter, appropriate restriction sites to facilitate cloning of inserts adjacent to the promoter, antibiotic resistance genes or other markers which  
15 can serve to select for cells containing the vector or the vector containing the insert, and necessary information processing sites, such as ribosome binding sites, RNA splice sites, polyadenylation sites and transcription termination sequences as well as any other sequence which may facilitate the expression of the inserted nucleic acid.

20 Preferred expression control sequences are promoters derived from metallothionine genes, actin genes, immunoglobulin genes, CMV, SV40, adenovirus, bovine papilloma virus, etc. A nucleic acid encoding a P/CAF protein can readily be determined based upon the genetic code for the amino acid sequence of the P/CAF protein and many nucleic acid sequences will encode a P/CAF protein. Modifications in  
25 the nucleic acid sequence encoding the P/CAF protein are also contemplated. Modifications that can be useful are modifications to the sequences controlling expression of the P/CAF protein to make production of P/CAF protein inducible or repressible as controlled by the appropriate inducer or repressor. Such means are standard in the art (*see, e.g., ref. 39*). The nucleic acids can be generated by means  
30 standard in the art, such as by recombinant nucleic acid techniques, as exemplified in the examples herein, and by synthetic nucleic acid synthesis or *in vitro* enzymatic synthesis.

After a nucleic acid encoding a particular P/CAF protein of interest, or a region of that nucleic acid, is constructed, modified, or isolated, that nucleic acid can then be cloned into an appropriate vector, which can direct the *in vivo* or *in vitro* synthesis of that wild-type and/or modified P/CAF protein. The vector is contemplated to have the  
5 necessary functional elements that direct and regulate transcription of the inserted nucleic acid, as described above. The vector containing the P/CAF nucleic acid or nucleic acid fragment can be in a host (e.g., cell or transgenic animal) for expressing the nucleic acid. The P/CAF protein or fragment thereof can thus be produced in a host system containing the expression vector and its functional activity as described herein  
10 can be demonstrated according to methods well known in the art.

There are numerous *E. coli* (*Escherichia coli*) expression vectors known to one of ordinary skill in the art useful for the expression of proteins. Other microbial hosts suitable for use include bacilli, such as *Bacillus subtilis*, and other enterobacteria, such  
15 as *Salmonella*, *Serratia*, as well as various *Pseudomonas* species. These prokaryotic hosts can support expression vectors which will typically contain expression control sequences compatible with the host cell (e.g., an origin of replication). In addition, any number of a variety of well-known promoters will be present, such as the lactose promoter system, a tryptophan (Trp) promoter system, a beta-lactamase promoter  
20 system, or a promoter system from phage lambda. The promoters will typically control expression, optionally with an operator sequence and have ribosome binding site sequences, for example, for initiating and completing transcription and translation. If necessary, an amino terminal methionine can be provided by insertion of a Met codon 5' and in-frame with the gene sequence. Also, the carboxy-terminal extension of the  
25 protein can be removed using standard oligonucleotide mutagenesis procedures.

Additionally, yeast expression can be used. There are several advantages to yeast expression systems. First, evidence exists that proteins produced in yeast secretion systems exhibit correct disulfide pairing. Second, post-translational glycosylation is  
30 efficiently carried out by yeast secretory systems. The *Saccharomyces cerevisiae* pre-pro-alpha-factor leader region (encoded by the *MF $\alpha$ -1* gene) is routinely used to direct

protein secretion from yeast (42). The leader region of pre-pro-alpha-factor contains a signal peptide and a pro-segment which includes a recognition sequence for a yeast protease encoded by the *KEX2* gene. This enzyme cleaves the precursor protein on the carboxyl side of a Lys-Arg dipeptide cleavage-signal sequence. The polypeptide coding sequence can be fused in-frame to the pre-pro-alpha-factor leader region. This construct is then put under the control of a strong transcription promoter, such as the alcohol dehydrogenase I promoter or a glycolytic promoter. The protein coding sequence is followed by a translation termination codon which is followed by transcription termination signals. Alternatively, the polypeptide encoding sequence of interest can be fused to a second protein coding sequence, such as Sj26 or  $\beta$ -galactosidase, used to facilitate purification of the resultant fusion protein by affinity chromatography. The insertion of protease cleavage sites to separate the components of the fusion protein is applicable to constructs used for expression in yeast.

Efficient post-translational glycosylation and expression of recombinant proteins can also be achieved in *Baculovirus* expression systems in insect cells.

Mammalian cells permit the expression of proteins in an environment that favors important post-translational modifications such as folding and cysteine pairing, addition of complex carbohydrate structures and secretion of active protein. Vectors useful for the expression of proteins in mammalian cells are characterized by insertion of the protein encoding sequence between a strong viral promoter and a polyadenylation signal. The vectors can contain genes conferring either gentamicin or methotrexate resistance for use as selectable markers. For example, the antigen and immunoreactive fragment coding sequence can be introduced into a Chinese hamster ovary (CHO) cell line using a methotrexate resistance-encoding vector. Presence of the vector RNA in transformed cells can be confirmed by Northern blot analysis and production of a cDNA or opposite strand RNA corresponding to the protein encoding sequence can be confirmed by Southern and Northern blot analysis, respectively. A number of other suitable host cell lines capable of secreting intact proteins have been developed in the art and include the CHO cell lines, HeLa cells, myeloma cell lines, Jurkat cells, and the like.

Expression vectors for these cells can include expression control sequences, as described above. The vectors containing the nucleic acid sequences of interest can be transferred into the host cell by well-known methods, which vary depending on the type of cell host. For example, calcium chloride transfection is commonly utilized for prokaryotic cells, whereas calcium phosphate treatment or electroporation may be used for other cell hosts.

Alternative vectors for the expression of protein in mammalian cells, similar to those developed for the expression of human gamma-interferon, tissue plasminogen activator, clotting Factor VIII, hepatitis B virus surface antigen, protease Nexin I, and eosinophil major basic protein, can be employed. Further, the vector can include CMV promoter sequences and a polyadenylation signal available for expression of inserted nucleic acid in mammalian cells (such as COS7).

The nucleic acid sequences can be expressed in hosts after the sequences have been positioned to ensure the functioning of an expression control sequence. These expression vectors are typically replicable in the host organisms either as episomes or as an integral part of the host chromosomal DNA. Commonly, expression vectors can contain selection markers, e.g., tetracycline resistance or hygromycin resistance, to permit detection and/or selection of those cells transformed with the desired nucleic acid sequences (*see*, e.g., U.S. Patent 4,704,362).

The nucleic acids produced as described above can also be expressed in a host which is a non-human animal to create a transgenic animal, containing, in a germ or somatic cell, a nucleic acid comprising the coding sequence for all or a portion of the P/CAF protein, as well as all of the other regulatory elements required for expression of the P/CAF protein-encoding sequence. The animal will express the P/CAF gene or portion thereof to produce the P/CAF protein or protein fragment and such expression can be detected by determination of a particular phenotype unique to the transgenic animal expressing the transferred nucleic acid.

The nucleic acid can be the nucleic acid of SEQ ID NO: 10, a nucleic acid having a nucleotide sequence which encodes the P/CAF protein, a nucleic acid having a nucleotide sequence which encodes the protein of SEQ ID NO: 1, as well as the nucleic acids that encode the proteins comprising the fragments of SEQ ID NOS: 2 and 4.

5

The nucleic acids of the invention can contain substitutions or deletions which provide a particular phenotype of interest. For example, various deletions or base substitutions can be introduced into the nucleic acid encoding the P/CAF protein for the purpose of studying the effects of these particular deletions or substitutions on the transcription modulation activity of the P/CAF protein. These effects can be monitored by observation of such characteristics as growth and development of the animal, the ability to develop tumors, survival rates and the like. The gene construct introduced into the animal cells to produce the transgenic animal can contain any of the regulatory elements described above to modulate expression of the foreign genes. As used herein, the term "phenotype" includes morphology, biochemical profiles, changes in tumor formation and other parameters that are affected by the presence of the P/CAF protein.

The transgenic animals of the invention can also be used in a method for determining the effectiveness of administering a nucleic acid encoding a functional P/CAF protein to a subject in need of a functional P/CAF protein. First, a nucleic acid encoding a nonfunctional P/CAF protein can be introduced into the animal's cells and expressed to yield a characteristic phenotype. Then, using standard gene therapy techniques, a nucleic acid encoding a functional P/CAF protein can be introduced into the animal's cells and the effects on the animal's phenotypic characteristics can be determined.

Having provided and taught how to obtain a nucleic acid that encodes a P/CAF protein, an isolated nucleic acid that encodes a fragment of P/CAF protein is also provided. The nucleic acid encoding the fragment can be obtained using any of the methods applicable to the nucleic acid encoding the entire P/CAF protein. The nucleic acid fragment can encode a species-specific P/CAF protein fragment (e.g., found in the

P/CAF protein of humans, but not in the P/CAF proteins of other species). Nucleic acids encoding species-specific fragments of P/CAF protein are themselves species-specific or allele-specific fragments of the P/CAF gene.

5           Examples of fragments of a nucleic acid encoding a fragment of the P/CAF protein can include the nucleic acid sequences which encode the amino acid sequences of the fragments of SEQ ID NOS:2 or 4. The same routine computer analyses used to select these examples of fragments can be routinely used to obtain others. Fragments of P/CAF-encoding nucleic acids can be primers for PCR or probes, which can be species-  
10 specific, gene-specific or allele-specific. P/CAF-encoding nucleic acid fragments can encode antigenic or immunogenic fragments of P/CAF protein that can be used in therapeutic assays or screening protocols. P/CAF gene fragments can encode fragments of P/CAF protein having histone acetylase activity and/or p300/CBP binding activity as described above, as well as other uses that may become apparent.

15

An isolated nucleic acid of at least ten nucleotides that selectively hybridizes with the nucleic acid of SEQ ID NO:10 under selected conditions is provided. For example, the conditions can be PCR amplification conditions and the hybridizing nucleic acid can be a primer consisting of a specific fragment of the reference sequence or a nearly  
20 identical nucleic acid that hybridizes only to the exemplified P/CAF-encoding nucleic acid or allelic variants thereof.

The invention provides an isolated nucleic acid that selectively hybridizes with the P/CAF-encoding nucleic acid sequence of SEQ ID NO:10 under stringent  
25 conditions. The hybridizing nucleic acid can be a probe that hybridizes only to the exemplified P/CAF-encoding nucleic acid sequence. Thus, the hybridizing nucleic acid can be a naturally occurring species-specific allelic variant of the exemplified P/CAF gene. The hybridizing nucleic acid can also include insubstantial base substitutions that do not prevent hybridization under the stated stringent conditions or affect either the  
30 function of the encoded protein, the way the protein accomplishes that function (e.g., its

secondary structure) or the ultimate result of the protein's activity. The means for determining these parameters are well known.

As used herein to describe nucleic acids, the term "selectively hybridizes" excludes the occasional randomly hybridizing nucleic acids as well as nucleic acids that encode other known homologs of the P/CAF protein. The selectively hybridizing nucleic acids of the invention can have at least 70%, 73%, 78%, 80%, 85%, 88%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% complementarity with the segment and strand of the sequence to which it hybridizes. This list is not intended to exclude percent complementarity values between these values. The nucleic acids can be at least 10, 15, 16, 17, 18, 20, 21, 23, 24, 25, 30, 35, 40, 50, 100, 150, 200, 300, 500, 550, 750, 900, 950, or 1000 nucleotides in length or any intervening length, depending on whether the nucleic acid is to be used as a primer, probe or for protein expression. The hybridizing nucleic acid can comprise a region of at least ten nucleotides (up to full length) that is completely complementary to a unique region of the nucleic acid to which it hybridizes.

The nucleic acid can be an alternative coding sequence for the P/CAF protein, or can be used as a probe or primer for detecting the presence of or obtaining the P/CAF protein. If used as primers, the invention provides compositions including at least two nucleic acids which selectively hybridize with different regions of the nucleic acid so as to amplify a desired region. Depending on the length of the probe or primer, it can range between 70% complementary bases and full complementarity and still hybridize under stringent conditions.

25

For example, for the purpose of obtaining or determining the presence of a nucleic acid encoding the P/CAF protein, the degree of complementarity between the hybridizing nucleic acid (probe or primer) and the sequence to which it hybridizes (P/CAF DNA in a sample) should be at least enough to exclude hybridization with a nucleic acid from another species. The invention provides examples of these nucleic acids of P/CAF, so that the degree of complementarity required to distinguish selectively

30

hybridizing from nonselectively hybridizing nucleic acids under stringent conditions can be clearly determined for each nucleic acid. It should also be clear that the hybridizing nucleic acids of the invention will not hybridize with nucleic acids encoding unrelated proteins (hybridization is selective) under stringent conditions.

5

"Stringent conditions" refers to the washing conditions used in a hybridization protocol. In general, the washing conditions should be a combination of temperature and salt concentration chosen so that the denaturation temperature is approximately 5-20°C below the calculated  $T_m$  of the nucleic acid hybrid under study. The temperature and salt conditions are readily determined empirically in preliminary experiments in which samples of reference DNA immobilized on filters are hybridized to the probe or protein encoding nucleic acid of interest and then washed under conditions of different stringencies. For example, the nucleic acid sequence of SEQ ID NO:10 was used as a specific radiolabeled probe for the detection of messenger RNA transcribed from the P/CAF gene by performing hybridizations under stringent conditions. The  $T_m$  of such an oligonucleotide can be estimated by allowing 2°C for each A or T nucleotide, and 4°C for each G or C. For example, an 18 nucleotide probe of 50% G+C would, therefore, have an approximate  $T_m$  of 54°C.

20 The invention provides an isolated nucleic acid that selectively hybridizes with the P/CAF gene shown in the sequence set forth as SEQ ID NO:10 under stringent conditions. The invention further provides an isolated nucleic acid complementary to the nucleotide sequence set forth in SEQ ID NO:10.

## 25 Antibodies to the P/CAF protein

A purified antibody and an antiserum containing polyclonal antibodies that specifically bind the P/CAF protein or antigenic fragment are also provided. The term "bind" means the well understood antigen/antibody binding as well as other nonrandom association with an antigen. "Specifically bind" as used herein describes an antibody or other ligand that does not cross react substantially with any antigen other than the one specified, in this case, an antigen of the P/CAF protein. Antibodies can be made as

30

described in Harlow and Lane (33). Briefly, purified P/CAF protein or an antigenic fragment thereof can be injected into an animal in an amount and in intervals sufficient to elicit a humoral immune response. Serum polyclonal antibodies can be purified directly, or spleen cells from the animal can be fused with an immortal cell line and screened for monoclonal antibody secretion, according to procedures well known in the art. Purified  
5 monospecific polyclonal antibodies that specifically bind the P/CAF antigen are also within the scope of the present invention. The antibodies of the present invention can bind the protein of claim 1, the protein of claim 2, the protein of claim 3 and/or the protein of claim 4, as well as any other proteins of the present invention.

10

A ligand that specifically binds the antigen is also contemplated. The ligand can be a fragment of an antibody, such as, for example, an Fab fragment which retains P/CAF binding activity, or a smaller molecule designed to bind an epitope of the P/CAF antigen. The antibody or ligand can be bound to a substrate or labeled with a detectable  
15 moiety or both bound and labeled. The detectable moieties contemplated within the compositions of the present invention include those listed above in the description of the diagnostic methods, including fluorescent, enzymatic and radioactive markers.

The antibody can be bound to a solid support substrate or conjugated with a  
20 detectable moiety or therapeutic compound or both bound and conjugated. Such conjugation techniques are well known in the art. For example, conjugation of fluorescent, radioactive or enzymatic moieties can be performed as described in the art (33,43). The detectable moieties contemplated in the present invention can include fluorescent, radioactive and enzymatic markers and the like. Therapeutic drugs  
25 contemplated with the present invention can include cytotoxic moieties such as ricin A chain, diphtheria toxin, pseudomonas exotoxin and other chemotherapeutic compounds.

It is well understood by one of skill in the art that all of the above discussion regarding antibodies to P/CAF can also be applied with regard to production,  
30 characterization and use of antibodies which bind the p300/CBP protein or any of the DNA-binding transcription factors of this invention.

### Measuring the P/CAF protein in a sample

The present invention also provides a method for determining the presence and thus the amount of P/CAF protein in a biological sample. As used herein, a biological sample includes any tissue or cell which would contain the P/CAF protein. Examples of  
5 cells include tissues taken from surgical biopsies, isolated from a body fluid or prepared in an *in vitro* tissue culture environment.

One example of determining the amount of P/CAF in a biological sample can comprise contacting the biological sample with a polypeptide comprising the amino acid  
10 sequence of SEQ ID NO:3 under conditions whereby a P/CAF/p300 complex can be formed; and determining the amount of the P/CAF/p300 complex, the amount of the complex indicating the amount of P/CAF in the sample. Determination of the amount of P/CAF/p300 complex can be accomplished through techniques standard in the art. For example, the complex may be precipitated out of a solution and detected by the  
15 addition of a detectable moiety conjugated to the p300 protein or by the detection of an antibody which binds p300 or the P/CAF protein, as taught in the Examples herein. Antibodies which bind p300 or the P/CAF protein can be either monoclonal or polyclonal antibodies and can be obtained as described herein. Detection of P/CAF/p300 complexes by the detection of the binding of antibodies reactive with p300  
20 or the P/CAF protein can be accomplished using various immunoassays as are available in the art, as described below.

Alternatively, determination of the amount of P/CAF in a biological sample can comprise contacting the biological sample with a polypeptide comprising the amino acid  
25 sequence of SEQ ID NO:9 under conditions whereby a P/CAF/CBP complex can be formed; and determining the amount of the P/CAF/CBP complex, the amount of the complex indicating the amount of P/CAF in the sample. Determination of the amount of P/CAF/CBP complex can be accomplished through techniques standard in the art. For example, the complex may be precipitated out of a solution and detected by the  
30 addition of a detectable moiety conjugated to the CBP protein or by the detection of an antibody which binds either CBP or the P/CAF protein, as taught in the Examples

herein. Antibodies which bind CBP or the P/CAF protein can be either monoclonal or polyclonal antibodies and can be obtained as described herein. Detection of P/CAF/CBP complexes by the detection of the binding of antibodies reactive with CBP or the P/CAF protein can be accomplished using various immunoassays as are available in the art, as described below.

Another example of determining the amount of P/CAF in a biological sample comprises contacting the biological sample with an antibody which specifically binds P/CAF under conditions whereby a P/CAF/ antibody complex can be formed and determining the amount of the P/CAF/antibody complex, the amount of the complex indicating the amount of P/CAF in the sample. Antibodies which bind P/CAF can be either monoclonal or polyclonal antibodies and can be obtained as described herein. Determination of P/CAF/antibody complexes can be accomplished using various immunoassays as are available in the art, as described below.

15

Immunoassays such as immunofluorescence assays, radioimmunoassays (RIA), immunoblotting and enzyme linked immunosorbent assays (ELISA) can be readily adapted for detection and measurement of P/CAF in a biological sample. Both polyclonal and monoclonal antibodies can be used in the assays. Available immunoassays are well known in the art and are extensively described in the patent scientific literature. See, for example, U.S. Patent Nos. 3,791,932; 3,839,153; 3,850,752; 3,850,578; 3,853,987; 3,867,517; 3,879,262; 3,901,654; 3,935,074; 3,984,533; 3,996,345; 4,034,074; and 4,098,876.

20

## 25 Screening assays for P/CAF

The present invention also provides a bioassay for screening substances for the ability to inhibit the histone acetyltransferase activity of P/CAF comprising contacting a system, in which histone acetylation by P/CAF can be determined, with the substance under conditions whereby histone acetylation by P/CAF can occur; determining the amount of histone acetylation by P/CAF in the presence of the substance; and comparing the amount of histone acetylation by P/CAF in the presence of the substance with the

30

amount of histone acetylation by P/CAF in the absence of the substance, a decreased amount of histone acetylation by P/CAF in the presence of the substance indicating a substance that can inhibit the histone acetyltransferase activity of P/CAF. The acetylation of histones by P/CAF can be determined in a system including, for example, either core histones (histones H2A, H2B, H3 and H4) or the nucleosome core particles (146 base pairs of DNA wrapped around the octamer of core histones) as substrates, the P/CAF protein and radiolabeled acetyl-CoA (e.g., [1-<sup>14</sup>C]acetyl CoA). The presence of acetylated histones can be detected by autoradiography after separation by SDS-PAGE as described herein in the Examples. Thus, the compound to be tested for the ability to inhibit the histone acetyltransferase activity of P/CAF can be added to this system and assayed for inhibiting ability.

The present invention also provides a bioassay for screening substances for the ability to inhibit the transcription modulating activity of P/CAF, comprising contacting a system, in which histone acetylation by P/CAF can be determined, with the substance under conditions whereby histone acetylation by P/CAF can occur; determining the amount of histone acetylation by P/CAF in the presence of the substance; and comparing the amount of histone acetylation by P/CAF in the presence of the substance with the amount of histone acetylation by P/CAF in the absence of the substance, a decreased amount of histone acetylation by P/CAF in the presence of the substance indicating a substance that can inhibit the transcription modulating activity and cell cycle progression suppressing activity of P/CAF. The acetylation of histones by P/CAF can be determined in a system including, for example, either core histones (histones H2A, H2B, H3 and H4) or the nucleosome core particles (146 base pairs of DNA wrapped around the octamer of core histones) as substrates, the P/CAF protein and radiolabeled acetyl-CoA (e.g., [1-<sup>14</sup>C]acetyl CoA). The presence of acetylated histones can be detected by autoradiography after separation by SDS-PAGE as described herein in the Examples. Thus, the compound to be tested for the ability to inhibit the transcription modulating activity of P/CAF by interfering with the histone acetyltransferase activity of P/CAF can be added to this system and assayed for inhibiting ability.

Also provided in the present invention is a bioassay for screening substances for the ability to inhibit the binding of p300 to P/CAF, comprising contacting a system in which the binding of p300 to P/CAF can be determined, with the substance under conditions whereby the binding of p300 and P/CAF can occur; determining the amount of p300 binding to P/CAF in the presence of the substance; and comparing the amount of p300 binding to P/CAF in the presence of the substance with the amount of p300 binding to P/CAF in the absence of the substance, a decreased amount of p300 binding to P/CAF in the presence of the substance indicating a substance that can inhibit the binding of p300 to P/CAF. The binding of p300 to P/CAF can be determined in a system, for example, which can include a cell free reaction mixture comprising a fragment of the p300 protein comprising the amino acid sequence of SEQ ID NO:3 and P/CAF. Alternatively, the system can comprise a cell extract produced from cells producing both p300 and P/CAF. Determination of the binding of p300 to P/CAF can be carried out as taught herein.

15

Additionally provided in the present invention is a bioassay for screening substances for the ability to inhibit the binding of CBP to P/CAF, comprising contacting a system in which the binding of CBP to P/CAF can be determined, with the substance under conditions whereby the binding of CBP to P/CAF can occur; determining the amount of CBP binding to P/CAF in the presence of the substance; and comparing the amount of CBP binding to P/CAF in the presence of the substance with the amount of CBP binding to P/CAF in the absence of the substance, a decreased amount of CBP binding to P/CAF in the presence of the substance indicating a substance that can inhibit the binding of CBP to P/CAF. The binding of CBP to P/CAF can be determined in a system, for example, which can include a cell free reaction mixture comprising a fragment of the CBP protein comprising the amino acid sequence of SEQ ID NO:9 and P/CAF. Alternatively, the system can comprise a cell extract produced from cells producing both CBP and P/CAF. Determination of the binding of CBP to P/CAF can be carried out as taught herein.

30

The present invention further contemplates a bioassay for screening substances for the ability to stimulate the histone acetyltransferase activity of P/CAF comprising contacting a system, in which histone acetylation by P/CAF can be determined, with the substance; determining the amount of histone acetylation by P/CAF in the presence of the substance; and comparing the amount of histone acetylation by P/CAF in the presence of the substance with the amount of histone acetylation by P/CAF in the absence of the substance, an increased amount of histone acetylation by P/CAF in the presence of the substance indicating a substance that can stimulate the histone acetyltransferase activity of P/CAF. The acetylation of histones by P/CAF can be determined in a system including, for example, either core histones (histones H2A, H2B, H3 and H4) or the nucleosome core particles (146 base pairs of DNA wrapped around the octamer of core histones) as substrates, the P/CAF protein and radiolabeled acetyl-CoA (e.g., [1-<sup>14</sup>C]acetyl CoA). The presence of acetylated histones can be detected by autoradiography after separation by SDS-PAGE as described herein in the Examples. Thus, the compound to be tested for the ability to stimulate the histone acetyltransferase activity of P/CAF can be added to this system and assayed for stimulating ability.

The present invention further contemplates a bioassay for screening substances for the ability to stimulate the transcription modulating activity of P/CAF comprising contacting a system, in which histone acetylation by P/CAF can be determined, with the substance; determining the amount of histone acetylation by P/CAF in the presence of the substance; and comparing the amount of histone acetylation by P/CAF in the presence of the substance with the amount of histone acetylation by P/CAF in the absence of the substance, an increased amount of histone acetylation by P/CAF in the presence of the substance indicating a substance that can stimulate the transcription modulating activity of P/CAF. The acetylation of histones by P/CAF can be determined in a system including, for example, either core histones (histones H2A, H2B, H3 and H4) or the nucleosome core particles (146 base pairs of DNA wrapped around the octamer of core histones) as substrates, the P/CAF protein and radiolabeled acetyl-CoA (e.g., [1-<sup>14</sup>C]acetyl CoA). The presence of acetylated histones can be detected by autoradiography after separation by SDS-PAGE as described herein in the Examples.

Thus, the compound to be tested for the ability to stimulate the transcription modulating activity of P/CAF by increasing the histone acetyltransferase activity of P/CAF can be added to this system and assayed for stimulating ability.

5           The present invention further provides a bioassay for screening substances for the ability to stimulate binding of p300 to P/CAF, comprising contacting a system in which the binding of p300 to P/CAF can be determined, with the substance under conditions whereby the binding of p300 to P/CAF can occur; determining the amount of p300 binding to P/CAF in the presence of the substance; and comparing the amount of  
10 p300 binding to P/CAF in the presence of the substance with the amount of p300 binding to P/CAF in the absence of the substance, an increased amount of p300 binding to P/CAF in the presence of the substance indicating a substance that can stimulate the binding of p300 to P/CAF. The binding of p300 to P/CAF can be determined in a system, for example, which can include a cell free reaction mixture comprising a  
15 fragment of the p300 protein comprising the amino acid sequence of SEQ ID NO:3 and P/CAF. Alternatively, the system can comprise a cell extract produced from cells producing both p300 and P/CAF. Determination of the binding of p300 to P/CAF can be carried out as taught herein.

20           Additionally provided in the present invention is a bioassay for screening substances for the ability to stimulate the binding of CBP to P/CAF, comprising contacting a system in which the binding of CBP to P/CAF can be determined, with the substance under conditions whereby the binding of CBP to P/CAF can occur; determining the amount of CBP binding to P/CAF in the presence of the substance; and  
25 comparing the amount of CBP binding to P/CAF in the presence of the substance with the amount of CBP binding to P/CAF in the absence of the substance, an increased amount of CBP binding to P/CAF in the presence of the substance indicating a substance that can stimulate the binding of CBP to P/CAF. The binding of CBP to P/CAF can be determined in a system, for example, which can include a cell free  
30 reaction mixture comprising a fragment of the CBP protein comprising the amino acid sequence of SEQ ID NO:9 and P/CAF. Alternatively, the system can comprise a cell

extract produced from cells producing both CBP and P/CAF. Determination of the binding of CBP to P/CAF can be carried out as taught herein.

#### **Transcription modulating activity of P/CAF**

5       The present invention contemplates a method for inhibiting the transcription modulating activity of P/CAF in a subject, comprising administering to the subject a transcription modulating activity inhibiting amount of a substance in a pharmaceutically acceptable carrier. For example, the substance can be identified according to the protocols provided herein as one that can inhibit the transcription modulating activity of  
10 P/CAF by preventing the binding of P/CAF to p300/CBP or by inhibiting the histone acetyltransferase activity of P/CAF as well as by any other inhibitory mechanism as identified by the protocols provided herein. Inhibition of the transcription modulating activity of P/CAF in a subject is desirable, for example, to inhibit HIV TAT-mediated transcription and therefore, the method of the present invention can be used to treat  
15 HIV-infected subjects.

The substance can be in a pharmaceutically acceptable carrier. By "pharmaceutically acceptable" is meant a material that is not biologically or otherwise undesirable, i.e., the material may be administered to a subject, along with the substance,  
20 without causing any undesirable biological effects or interacting in a deleterious manner with any of the other components of the pharmaceutical composition in which it is contained. The carrier would naturally be selected to minimize any degradation of the active ingredient and to minimize any adverse side effects in the subject.

25       The transcription modulating activity and/or histone acetyltransferase activity of P/CAF can be inhibited in a subject by administering to the subject a substance which binds p300/CBP at the P/CAF binding site or a substance which binds the P/CAF protein at the p300/CBP binding site, the ultimate result being that P/CAF and p300/CBP do not bind with one another and P/CAF cannot exert its transcription  
30 modulating and/or histone acetyltransferase effect. The substance can be a protein, such as an antibody which binds the P/CAF protein binding site at or near the p300/CBP

binding site, thereby preventing its binding or an antibody which binds the p300/CBP protein at or near the P/CAF binding site, thereby preventing its binding. The substance can also bind the histone acetyltransferase site on P/CAF or at the acetylation site on the histone, thereby preventing acetylation by P/CAF.

5

The substance which binds p300/CBP, the P/CAF protein or the histone and has the net effect of inhibiting the transcription modulating effect and or histone acetyltransferase activity of P/CAF in the cell can be delivered to a cell in the subject by mechanisms well known in the art.

10

Alternatively, a nucleic acid encoding a protein which binds either to p300/CBP or the P/CAF protein and has the net effect of inhibiting the transcription modulating effect and/or histone acetyltransferase activity of P/CAF in the cell can be delivered to a cell in the subject by gene transduction mechanisms well known in the art. For example, nucleic acid can be introduced by liposomes as well as via retroviral or adeno-associated viral vectors, as described below.

15

The substance which inhibits the transcription modulating effect and/or histone acetyltransferase activity of P/CAF can be an antisense RNA or an antisense DNA which binds the RNA or DNA of P/CAF, thereby preventing translation or transcription of the RNA or DNA encoding P/CAF and having the net effect of inhibiting the transcription modulating effect and/or histone acetyltransferase activity of P/CAF by inhibiting P/CAF production. The antisense RNA of the present invention can be generated from the nucleic acid of SEQ ID NO:14 (human) or SEQ ID NO:15 (mouse). Furthermore, the antisense DNA can be a phosphorothioate oligodeoxynucleotide having the nucleotide sequence of SEQ ID NO:16 (human) or of SEQ ID NO:17 (mouse). The mouse antisense RNA can be used to inhibit the activity of mouse P/CAF, having the nucleotide sequence of SEQ ID NO:18 and the amino acid sequence of SEQ ID NO:8. The present invention also contemplates an antisense nucleic acid sequence which can bind the DNA or RNA of any of the transcription factors or other proteins now known or later identified to bind P/CAF, thereby inhibiting expression of the gene products of

20  
25  
30

these proteins and having the net effect of inhibiting the transcription modulating effect and/or histone acetyltransferase activity of P/CAF.

The antisense nucleic acid can comprise a typical nucleic acid, but the antisense  
5 nucleic acid can also be a modified nucleic acid or a derivative of a nucleic acid such as a phosphorothioate analogue of a nucleic acid. The composition can comprise, for example, an antisense RNA that specifically binds an RNA encoded by the gene encoding the serum protein. Antisense RNAs can be synthesized and used by standard methods (62).

10

Antisense RNA can inhibit gene expression by forming an RNA/RNA duplex  
between the antisense RNA and the RNA transcribed from the target gene. The precise  
mechanism by which this duplex formation decreases the production of the protein  
encoded by the endogenous gene probably involves binding of complementary regions  
15 of the normal sense mRNA and the antisense RNA strand with duplex formation in a manner that blocks RNA processing and translation. Alternative mechanisms include the formation of a triplex between the antisense RNA and duplex DNA or the formation of an DNA-RNA duplex with subsequent degradation of DNA-RNA hybrids by RNase H. Furthermore, an antigene effect can result from certain DNA-based oligonucleotides  
20 via triple-helix formation between the oligomer and double-stranded DNA which results in the repression of gene transcription. Regardless of the specific molecular mechanism, the present invention results in inhibition of expression of the P/CAF gene by the introduced and replicated DNA resulting in inhibition of the transcription modulating and/or histone acetyltransferase activity of P/CAF, by a reduction in the expression of  
25 the nucleic acid to which the antisense nucleic acid is hybridized, and therefore a reduction of the gene product from the targeted gene.

The antisense nucleic acid may be obtained by any number of techniques known  
to one skilled in the art. One method of constructing an antisense nucleic acid is to  
30 synthesize a recombinant antisense DNA molecule. For example, oligonucleotide synthesis procedures are routine in the art and oligonucleotides coding for a particular

protein or regulatory region are readily obtainable through automated DNA synthesis. A nucleic acid for one strand of a double-stranded molecule can be synthesized and hybridized to its complementary strand. One can design these oligonucleotides such that the resulting double-stranded molecule has either internal restriction sites or appropriate 5' or 3' overhangs at the termini for cloning into an appropriate vector. Double-stranded molecules coding for relatively large proteins or regulatory regions can be synthesized by first constructing several different double-stranded molecules that code for particular regions of the protein or regulatory region, followed by ligating these DNA molecules together. Once the appropriate DNA molecule is synthesized, this DNA can be cloned downstream of a promoter in an antisense orientation. Techniques such as this are routine in the art and are well documented.

An example of another method of obtaining an antisense nucleic acid is to isolate that nucleic acid from the organism in which it is found and clone it in an antisense orientation. For example, a DNA or cDNA library can be constructed and screened for the presence of the nucleic acid of interest. Methods of constructing and screening such libraries are well known in the art and kits for performing the construction and screening steps are commercially available (for example, Stratagene Cloning Systems, La Jolla, CA). Once isolated, the nucleic acid can be directly cloned into an appropriate vector in an antisense orientation, or if necessary, be modified to facilitate the subsequent cloning steps. Such modification steps are routine, an example of which is the addition of oligonucleotide linkers which contain restriction sites to the termini of the nucleic acid. General methods are set forth in Sambrook *et al.* (39).

The DNA that is introduced into the cell is in an expression orientation that is antisense to a corresponding endogenous DNA or RNA of the cells. For example, where an endogenous DNA comprises a gene which encodes for a particular protein, the introduced DNA is in an expression orientation opposite the expression of the endogenous DNA; that is the DNA operatively linked to a promoter is in an antisense expression orientation relative to the corresponding endogenous gene. The introduced DNA may be homologous to the entire transcribed gene or homologous to only part of

the transcribed gene. Alternatively, the sequence of the introduced DNA may be divergent to that of the endogenous DNA but only divergent to the extent that hybridization of the nucleic acids occurs, thereby preventing transcription. One skilled in the art can determine the maximum extent of this divergence by routine screening of antisense DNAs corresponding to an endogenous DNA of the cell. In this manner, one skilled in the art can readily determine which fragments, or alternatively the extent of homology of the fragments or the entire gene that is necessary to inhibit gene expression.

10           The antisense nucleic acids of the present invention can be made according to protocols standard in the art, as well as described in the Examples provided herein. The antisense nucleic acids can be administered to a subject according to the gene transduction protocols standard in the art, as described below.

15           The present invention also contemplates a method for stimulating the transcription modulating activity and/or histone acetyltransferase activity of P/CAF in a subject comprising administering to the subject a substance, in a pharmaceutically acceptable carrier, determined according to the methods taught herein, to have a stimulatory affect on the transcription modulating and/or histone acetyltransferase activity of P/CAF. The substance can be one which has been identified, according to the protocols provided herein, to stimulate histone acetyltransferase activity in P/CAF or promote binding of P/CAF to p300/CBP. The stimulation of the transcription modulation activity and/or histone acetyltransferase activity of P/CAF in a subject is desirable, for example, to activate tumor suppressor p53 (which promotes apoptosis) or to activate the muscle differentiation factor, MyoD. Thus, the method of the present invention can be employed to treat cancer and to promote muscle differentiation in conditions where muscle differentiation is desired. The substance can be delivered to a cell in the subject by mechanisms well known in the art.

30           Further contemplated in the present invention is a method for promoting binding of P/CAF to p300/CBP in a subject, comprising administering to the subject a substance

identified by the methods provided herein to promote binding of P/CAF to either p300 or CBP.

Additionally, a nucleic acid encoding a protein which stimulates the transcription  
5 modulating activity and/or histone acetyltransferase activity of P/CAF can be delivered to a cell in the subject by gene transduction mechanisms, as described below.

Also provided in the present invention is a method of inhibiting the cell cycle  
progression inducing effect of an oncoprotein which binds p300/CBP in a subject  
10 comprising transducing the cells of the subject with a vector comprising a nucleic acid encoding the P/CAF protein; inducing expression of the nucleic acid in the cell to produce the P/CAF in an amount which will allow the P/CAF gene product to replace the oncoprotein bound to p300/CBP, whereby the replacement of the oncoprotein  
bound to p300/CBP by the P/CAF gene product inhibits the cell cycle progression  
15 inducing effect of the oncoprotein. The oncoprotein which binds p300/CBP in the cell can be the adenovirus E1A oncoprotein.

A method for providing a functional P/CAF protein to a subject in need of the  
functional P/CAF protein is also provided, comprising transducing the cells of the  
20 subject with a vector comprising a nucleic acid encoding the P/CAF protein and inducing expression of the nucleic acid to produce the functional P/CAF protein in the cell, thereby providing the functional P/CAF protein to the subject. The transduction of the vector nucleic acid into the subject's cells can be carried out according to standard gene therapy protocols well known in the art (see, for example, U.S. Patent No.  
25 5,339,346).

#### **Screening assays for p300/CBP**

The present invention also provides a bioassay for screening substances for the  
ability to inhibit the histone acetyltransferase activity of p300/CBP comprising  
30 contacting a system, in which histone acetylation by p300/CBP can be determined, with the substance under conditions whereby histone acetylation by p300/CBP can occur;

determining the amount of histone acetylation by p300/CBP in the presence of the substance; and comparing the amount of histone acetylation by p300/CBP in the presence of the substance with the amount of histone acetylation by p300/CBP in the absence of the substance, a decreased amount of histone acetylation by p300/CBP in the presence of the substance indicating a substance that can inhibit the histone acetyltransferase activity of p300/CBP. The acetylation of histones by p300/CBP can be determined in a system including, for example, either core histones (histones H2A, H2B, H3 and H4) or the nucleosome core particles (146 base pairs of DNA wrapped around the octamer of core histones) as substrates, the P300/CBP protein and radiolabeled acetyl-CoA (e.g., [1-<sup>14</sup>C]acetyl CoA). The presence of acetylated histones can be detected by autoradiography after separation by SDS-PAGE as described herein in the Examples. Thus, the compound to be tested for the ability to inhibit the histone acetyltransferase activity of p300/CBP can be added to this system and assayed for acetyltransferase inhibiting ability.

15

Also provided in the present invention is a bioassay for screening substances for the ability to inhibit the binding of a transcriptional factor to p300/CBP, comprising contacting a system in which the binding of a transcriptional factor to p300/CBP can be determined, with the substance under conditions whereby the binding of the transcriptional factor and p300/CBP can occur; determining the amount of transcriptional factor binding to p300/CBP in the presence of the substance; and comparing the amount of transcriptional factor binding to p300/CBP in the presence of the substance with the amount of transcriptional factor binding to p300/CBP in the absence of the substance, a decreased amount of transcriptional factor binding to p300/CBP in the presence of the substance indicating a substance that can inhibit the binding of a transcriptional factor to p300/CBP. The binding of a transcriptional factor to p300/CBP can be determined in a system, for example, which can include a cell free reaction mixture comprising a transcriptional factor which binds p300/CBP and p300/CBP. Alternatively, the system can comprise a cell extract produced from cells producing both a transcriptional factor which binds p300/CBP and p300/CBP. The transcriptional factor which binds p300/CBP can be selected from, but is not limited to

30

the group consisting of nuclear hormone receptors, CREB, c-Jun/v-Jun, c-Myb/v-Myb, YY1, Sap-1a, c-Fos, MyoD and SRC-1, as well as any other transcriptional factor now known or later identified to bind p300/CBP. The screening assay of the present invention can also be used to identify substances which inhibit the binding of p300/CBP to other components to which it is known to bind, for example, P/CAF, pp90<sub>RSK</sub>, TFIIB, E1A, SV40 large T antigen, as well as any other substances now known or later identified to bind p300/CBP. Determination of the binding of a transcriptional factor or other substance to p300/CBP can be carried out as taught in the Examples herein as well as by protocols described in the literature.

10

The present invention further contemplates a bioassay for screening substances for the ability to stimulate the histone acetyltransferase activity of p300/CBP comprising contacting a system, in which histone acetylation by p300/CBP can be determined, with the substance; determining the amount of histone acetylation by p300/CBP in the presence of the substance; and comparing the amount of histone acetylation by p300/CBP in the presence of the substance with the amount of histone acetylation by p300/CBP in the absence of the substance, an increased amount of histone acetylation by p300/CBP in the presence of the substance indicating a substance that can stimulate the histone acetyltransferase activity of p300/CBP. The acetylation of histones by p300/CBP can be determined in a system including, for example, either core histones (histones H2A, H2B, H3 and H4) or the nucleosome core particles (146 base pairs of DNA wrapped around the octamer of core histones) as substrates, the p300/CBP protein and radiolabeled acetyl-CoA (e.g., [1-<sup>14</sup>C]acetyl CoA). The presence of acetylated histones can be detected by autoradiography after separation by SDS-PAGE as described herein in the Examples. Thus, the compound to be tested for the ability to stimulate the histone acetyltransferase activity of p300/CBP can be added to this system and assayed for stimulating ability.

The present invention further provides a bioassay for screening substances for the ability to stimulate binding of a component, which binds p300/CBP, to p300/CBP, comprising contacting a system in which the binding of the component to p300/CBP can

be determined, with the substance under conditions whereby the binding of the component to p300/CBP can occur; determining the amount of component binding to p300/CBP in the presence of the substance; and comparing the amount of component binding to p300/CBP in the presence of the substance with the amount of component  
5 binding to p300/CBP in the absence of the substance, an increased amount of component binding to p300/CBP in the presence of the substance indicating a substance that can stimulate the binding of the component to p300/CBP. The binding of the component to p300/CBP can be determined in a system, for example, which can include a cell free reaction mixture comprising the component and p300/CBP. Alternatively, the  
10 system can comprise a cell extract produced from cells producing both the component and p300/CBP. The component which binds p300/CBP can be any of the transcriptional factors or other proteins which are known or are identified in the future to bind p300/CBP, as set forth above. Determination of the binding of the component to p300/CBP can be carried out as taught in the Examples provided herein and according  
15 to protocols available in the literature.

#### **Histone acetyltransferase activity of p300/CBP**

A method for inhibiting the histone acetyltransferase activity of p300/CBP in a subject is provided in the present invention, comprising administering to the subject a  
20 histone acetyltransferase activity inhibiting amount of a substance in a pharmaceutically acceptable carrier. The mechanism of the inhibitory action of the substance can be the inhibition of the binding of a DNA-binding transcription factor, such as, for example, a nuclear hormone receptor, CREB, c-Jun/v-Jun, c-Myb/v-Myb, YY1, Sap-1a, c-Fos, MyoD or SRC-1, to p300/CBP.

25

The histone acetyltransferase activity of p300/CBP can be inhibited in a subject by administering to the subject a substance which binds p300/CBP at the transcription factor binding site or a substance which binds the transcription factor protein at the p300/CBP binding site, the ultimate result being that the transcription factor and  
30 p300/CBP do not bind with one another and p300/CBP cannot acetylate histones.

The substance which binds either to the transcription factor or the p300/CBP protein and has the net effect of inhibiting the histone acetyltransferase activity of p300/CBP in the cell can be identified according to the screening methods provided herein and delivered to a cell in the subject by mechanisms well known in the art. The substance can be a protein, such as an antibody which binds the p300/CBP protein binding site at or near the DNA-binding transcription factor binding site, thereby preventing its binding or an antibody which binds the DNA-binding transcription factor at or near the p300/CBP binding site, thereby preventing its binding. The substance can also bind the histone acetyltransferase site on p300/CBP (aa 1195-1673 on p300 or aa 1174-1850 on CBP) or at the acetylation site on the histone, thereby preventing acetylation by p300/CBP.

Additionally, the substance can be a nucleic acid which can be expressed in the cell to produce a protein which inhibits the histone acetyltransferase activity of p300/CBP. For example, a nucleic acid encoding a protein which binds either to a transcription factor or the p300/CBP protein and has the net effect of inhibiting the histone acetyltransferase activity of p300/CBP in the cell can be delivered to a cell in the subject by gene transduction mechanisms well known in the art. For example, nucleic acid can be introduced by liposomes as well as via retroviral or adeno-associated viral vectors, as described below.

The substance which inhibits the histone acetyltransferase activity of p300/CBP can be an antisense RNA or an antisense DNA which binds the RNA or DNA of p300/CBP thereby preventing translation or transcription of the RNA or DNA encoding p300/CBP and having the net effect of inhibiting the histone acetyltransferase activity of P/CAF by inhibiting p300/CBP production. The antisense RNA or DNA of the present invention can be produced and introduced into cells according to the same methods as set forth above for P/CAF antisense nucleic acids.

The present invention also contemplates a method for stimulating the histone acetyltransferase activity of p300/CBP in a subject comprising administering to the

subject a histone acetyltransferase activity stimulating amount of a substance, in a pharmaceutically acceptable carrier, determined according to the methods taught herein, to have a stimulatory affect on the histone acetyltransferase activity of p300/CBP. The substance can exert a stimulatory effect by promoting the binding of a DNA-binding transcription factor of the present invention to p300/CBP. The substance can be delivered to a cell in the subject by mechanisms well known in the art. A nucleic acid encoding a protein which stimulates the transcription modulating activity of p300/CBP can be delivered to a cell in the subject by gene transduction mechanisms, as described below.

#### **Gene transduction**

In the methods described above which include gene transduction into cells (i.e., addition of exogenous DNA into cells), the nucleic acids of the present invention can be in a vector for delivering the nucleic acids to the site for expression of the P/CAF protein. The vector can be one of the commercially available preparations, such as the pGM plasmid (Promega). Vector delivery can be by liposome, using commercially available liposome preparations or newly developed liposomes having the features of the present liposomes. Additionally, vector delivery can be via a viral system, including, but not limited to, retroviral, adenoviral and adeno-associated viral systems. Other delivery methods can be adopted and routinely tested according to the methods taught herein.

The modes of administration of the liposome will vary predictably according to the disease being treated and the tissue being targeted. For example, for treating cancer in either the lung or the liver, which are both sinks for liposomes, intravenous delivery is reasonable. For other localized cancers, as well as precancerous conditions, catheterization of an artery upstream from the target organ is a preferred mode of delivery, because it avoids significant clearance of the liposome by the lung and liver. For cancerous lesions at a number of other sites (e.g., skin cancer, localized dysplasias), topical delivery is expected to be effective and may be preferred, because of its convenience.

Leukemias and other disorders involving dysregulated proliferation of certain isolatable cell populations may be more readily treated by *ex vivo* administration of the nucleic acid.

5           The liposomes may be administered topically, parenterally (e.g., intravenously), by intramuscular injection, by intraperitoneal injection, transdermally, extracorporeally or the like, although intravenous or topical administration is typically preferred. The exact amount of the liposomes required will vary from subject to subject, depending on the species, age, weight and general condition of the subject, the severity of the disease  
10   being treated, the particular compound used, its mode of administration and the like. Thus, it is not possible to specify an exact amount. However, an appropriate amount may be determined by one of ordinary skill in the art using only routine experimentation given the teachings herein.

15           Parenteral administration, if used, is generally characterized by injection. Injectables can be prepared in conventional forms, either as liquid solutions or suspensions, solid forms suitable for solution or suspension in liquid prior to injection, or as emulsions. A more recently revised approach for parenteral administration involves use of a slow release or sustained release system such that a constant level of dosage is  
20   maintained. See, e.g., U.S. Patent No. 3,610,795, which is incorporated by reference herein.

Topical administration can be by creams, gels, suppositories and the like. *Ex vivo* (extracorporeal) delivery can be as typically used in other contexts.

25

The present invention is more particularly described in the following examples which are intended as illustrative only since numerous modifications and variations therein will be apparent to those skilled in the art.

30

## EXAMPLES

### I. P/CAF studies.

#### 5 Cloning and characterization of P/CAF protein.

In human cells, CBP binds to c-Jun in a phosphorylation-dependent manner in association with stimulation of transcription (9). In yeast, GCN4 is believed to be a c-Jun counterpart on the basis of similarities in DNA recognition (15) as well as the participation of both proteins in UV signaling pathways (16). Yeast genetic screening  
10 has led to the isolation of various cofactors for GCN4, including GCN5 (yGCN5), ADA2 (yADA2) and ADA3 (yADA3) (17-19). These factors are considered to function as a complex (or in a common pathway) based on genetic and protein-protein interaction studies (18-22). Finally, p300/CBP and yADA2 exhibit significant sequence similarity within a 50 amino acid region including a Zn<sup>2+</sup> finger motif (3). Human  
15 counterparts to yGCN5, yADA2, or yADA3 that interact with p300/CBP to mediate transcriptional activation by c-Jun were searched for in various nucleotide sequence databases.

Comparison of the yGCN5 protein sequence with various databases (23)  
20 revealed significant similarities with the two randomly sequenced human cDNAs, ETS05039 (24) ( $P=4.0 \times 10^{-15}$ ) and NIB2000-5R ( $P=6.5 \times 10^{-9}$ ). Given that these cDNAs were truncated, human fetal liver and fetal brain cDNA libraries (Clontech) were screened with ETS05039 and NIB2000-5R, respectively and complete clones were isolated from the human fetal liver cDNA library. The complete sequences revealed that  
25 the ETS05039- and NIB2000-5R-derived clones are encoded by distinct genes but are highly related within the protein coding regions (68% identity at the DNA level; 75% identity and 86% similarity at the protein level). The former encodes an N-terminal region with no sequence similarity to any proteins in the databases besides the yGCN5-related C-terminal region, whereas the latter encodes only the yGCN5-related region.  
30 Given that p300/CBP-binding activity was observed in the former polypeptide as shown below, it was designated p300/CBP-associated factor (P/CAF), having the amino acid

sequence of SEQ ID NO:1 and the nucleotide sequence of SEQ ID NO:10 and the latter was named human GCN5 (hGCN5), having the amino acid sequence of SEQ ID NO:5 and the nucleotide sequence of SEQ ID NO:11.

5           Additionally, an RNA blot (Clontech) was hybridized with a random-primed probe made from the cDNA encoding P/CAF. RNA blotting indicated that transcripts detected by the P/CAF and hGCN5 cDNAs are ubiquitously expressed, but the former is most abundant in heart and skeletal muscle, whereas the latter is most abundant in pancreas and skeletal muscle.

10

#### **P/CAF-p300/CBP interaction *in vitro***

          The P/CAF binding site was presumed to reside in the C terminal one third of CBP (residues 1,678-2,442) because it was observed that this region, when fused to a DNA binding domain, activates transcription (4) in a manner repressed by coexpression  
15 of 12S E1A. This region was divided into 6 overlapping fragments and each was expressed in *E. coli* as a glutathione-S-transferase (GST) fusion protein. GST-CBP fusions were incubated with recombinant P/CAF protein and, subsequently, purified using glutathione-Sepharose. Co-purified P/CAF was detected by immunoblotting analysis.

20

          To construct GST-fusions, various regions of CBP and p300 were amplified by PCR. A series of deletions of the CBP segment B was created by site-directed *in vitro* mutagenesis (30). These fragments were subcloned into pGEX-2T (Pharmacia). GST-fusions were expressed in *E. coli* and extracted with buffer B [20 mM Tris-HCl (pH  
25 8.0), 5 mM MgCl<sub>2</sub>, 10% glycerol, 1 mM AEBSF, 0.1% NP40, 10 µg/ml of aprotinin, 10 µg/ml of leupeptin, 1 µg/ml of pepstatin A, 1 mM DTT] containing 0.1 M KCl for these experiments. GST-CBP-segment B was purified by glutathione-Sepharose and phenyl-Sepharose chromatographic steps, P/CAF, hGCN5, and E1A were expressed as FLAG-fusions in Sf9 cells via baculovirus vectors and affinity-purified with M2-agarose (ref.  
30 30; Kodak-IBI). For interaction, a crude *E. coli* extract containing 20 pmol of GST-fusion was incubated with 40-60 pmol of P/CAF or E1A in a total volume of 50 µl of

buffer B with 0.1 M KCl on ice for 10 min. Samples were further incubated with 10  $\mu$ l (packed volume) of glutathione-Sepharose at 4°C for 30 min, washed four times with 200  $\mu$ l of buffer B containing 0.1 M KCl, and eluted with 20  $\mu$ l of buffer E [50 mM Tris-HCl (pH 8.0), 0.2 M KCl, 20 mM glutathione] for 60 min. Interacting proteins  
5 were detected by anti-FLAG immunoblotting or silver staining.

For p300 interactions, the segment spanning residues 1763-1966 (segment B') of p300, which is analogous to the CBP segment-B, was used. Twenty percent of the P/CAF and hGCN5 inputs and 100% of the E1A input were also analyzed. In the GST  
10 precipitation assays, almost identical amounts of the GST fusions were recovered in all samples. Interaction between P/CAF and CBP (segment B) was determined in the absence and in the presence of E1A. Control reactions with GST-CBP alone and without GST-CBP were also performed. Input proteins were analyzed.

15 Two CBP segments, A and B, interacted specifically with P/CAF. The stronger interaction was observed in the latter segment, which does not include the yADA2-like Zn<sup>2+</sup> finger. Given that the CBP segment-B is well conserved in p300 (66% identity, 75% similarity), the binding of P/CAF to p300 *in vitro* was also analyzed. For this experiment, the p300 segment spanning residues 1763-1966, termed segment B', which  
20 is analogous to the CBP segment-B, was used. Like CBP, p300 interacted specifically with P/CAF. These studies demonstrated that P/CAF binds specifically to both p300 and CBP *in vitro*. In contrast to P/CAF, hGCN5 did not bind to CBP or p300.

These studies also demonstrated that the Zn<sup>2+</sup> finger region of p300/CBP, which  
25 shares sequence similarity with yADA2, is not essential for the interaction with P/CAF. Cloning of a human structural homolog of yADA2, termed hADA2 (25) has revealed that, unlike the sequence similarity between p300/CBP and yADA2, which is restricted to a 50 amino acid region, hADA2 shares extensive similarity (30% identity, 52% similarity) to yADA2 over the entire protein sequence. Moreover, a computer search of  
30 the complete genomic sequence of *Saccharomyces cerevisiae* revealed that yeast does

not have counterparts of p300/CBP or P/CAF. Thus, the p300/CBP-P/CAF pathway may have been acquired during metazoan evolution.

## 5 Action of E1A *in vitro*

Previous reports indicated that E1A binds to both the p300 segment spanning residues 1767-1816 and the CBP segment spanning residues 1805-1854 (7). These interactions were reconfirmed in the present system; thus, both p300 and CBP segments covering the previously identified regions interacted with E1A.

10

For further mapping, a series of deletions was introduced within the CBP segment-B and tested for interactions with P/CAF and E1A. Deletions of residues 1801-1825 or 1824-1851 markedly reduced interactions with both P/CAF and E1A, whereas deletion of residues 1850-1878 did not affect these interactions. Furthermore, deletion of residues 1801-1851 completely abolished interactions with both P/CAF and E1A. These data indicate that residues 1801-1851 of CBP are critical for interaction with both P/CAF and E1A. Taken together with the evidence that CBP segment A (aa residues 1,678-1,880) also binds to these factors, the above findings demonstrate that P/CAF and E1A bind to the same or very closely spaced sites on CBP.

20

Evidence that both P/CAF and E1A recognize the same p300/CBP segments raises the possibility of direct competition between P/CAF and E1A for binding to p300/CBP. To test this possibility, a competition experiment was performed with the use of affinity purified recombinant proteins. The interaction of P/CAF with the CBP-segment B was progressively inhibited by the addition of increasing amounts of E1A. In contrast, no inhibition was caused by an E1A mutant which does not bind to p300/CBP (E1A $\Delta$ N). Similar results were obtained with the p300-segment B', leading to the conclusion that P/CAF and E1A compete for the same binding sites in p300/CBP.

30

**P/CAF-p300/CBP interaction *in vivo***

The *in vivo* interaction between P/CAF and p300/CBP was established by co-immunoprecipitation from a human osteosarcoma cell extract. Proteins in this extract were immunoprecipitated with rabbit anti-P/CAF, rabbit anti-CBP and anti-p300  
5 antibodies. For controls, cell extract was precipitated with rabbit control IgG or mouse anti-HA monoclonal antibody. The precipitates were analyzed by immunoblotting with anti-P/CAF, anti-CBP and anti-p300 antibodies.

Osteosarcoma cells were transfected with either control vector or E1A- or  
10 E1AΔN-expression vectors. Extract from the transfected subpopulation was immunoprecipitated with anti-P/CAF or control IgG. The precipitates were analyzed by immunoblotting with anti-p300 and anti-P/CAF antibodies.

Rabbit anti-P/CAF antibody was raised to the P/CAF segment spanning residues  
15 125-397 and purified by immunoaffinity chromatography (33). A mixture of monoclonal antibodies raised to the human p300 segment spanning residues 1572-2371 (5) and rabbit polyclonal antibodies raised to the mouse CBP segment spanning residues 2-23 (for immunoprecipitation) and 1736-2179 (immunoblotting) were purchased from Upstate Biotechnology. Approximately  $2 \times 10^7$  human osteosarcoma U-2 OS cells  
20 (ATCC accession number HTB 96) were extracted with 10 ml of lysis buffer [25 mM HEPES-KOH (pH 7.2), 150 mM potassium acetate, 2 mM EDTA, 1 mM DTT, 1 mM AEBSF, 10 µg/ml of aprotinin, 10 µg/ml of leupeptin, 1 µg/ml of pepstatin A, 20 mM sodium fluoride, 0.1% NP40]. Two to 10 ml of extract were incubated with 2 µg of the respective antibody for four hours at 4°C. Fifty µl (packed volume) of protein-A  
25 Trisacryl (Pierce) were added and incubation was continued for two hours. The matrix was washed four times with 1 ml of the lysis buffer, then boiled in 2x SDS sample buffer. Human osteosarcoma U-2 OS cells were transfected with 20 µg of the indicated plasmid and 1 µg of sorting plasmid (pCMV-IL2R) (31). The transfected subpopulation was purified by magnetic affinity cell sorting (32). Extract from approximately  $2 \times 10^5$   
30 sorted cells was immunoprecipitated as described.

Anti-P/CAF antibody specifically detected a 95 kDa protein, which is very close to the calculated value for the full-length P/CAF, in the immunoprecipitates. Anti-P/CAF antibody co-immunoprecipitated both CBP and p300. Similarly, anti-CBP antibody also co-immunoprecipitated P/CAF. However, anti-p300 antibody did not co-immunoprecipitate P/CAF. This is most likely due to steric interference since the anti-p300 antibody was raised to the p300 segment spanning residues 1572-2371 which includes the P/CAF binding region. These data demonstrate that P/CAF forms complexes with both p300 and CBP *in vivo*.

#### 10 Action of E1A *in vivo*

The *in vitro* experiments described herein indicate that P/CAF and E1A compete for the binding sites in p300/CBP. Thus, a study was conducted to determine whether E1A targets the endogenous interaction between P/CAF and p300. An E1A-expression vector was transiently transfected into human osteosarcoma cells and the transfected subpopulation was purified by cell sorting. Then, the interaction between P/CAF and p300 in transfected cells was examined by co-immunoprecipitation with anti-P/CAF antibody. The endogenous interaction of P/CAF with p300 was drastically inhibited by expression of E1A. On the other hand, no inhibition was observed by the E1A mutant lacking the p300 binding domain (E1A $\Delta$ N), indicating that E1A disrupts the P/CAF-p300 complex *in vivo* through an interaction with p300.

#### Cell cycle regulation by P/CAF

Given that binding of P/CAF to p300/CBP is inhibited by E1A, experiments were performed to evaluate whether P/CAF, by binding to and forming a functional complex with p300, is involved in the regulation of entry into S phase. This possibility was addressed by examining whether transient expression of P/CAF would affect the rate of G1/S transit in HeLa cells. P/CAF negatively affected the distribution of cells between G1 and S phases in this assay.

30 HeLa cells were transfected by electroporation with 7  $\mu$ g of P/CAF-expression plasmid and/or 3  $\mu$ g of the full-length or the N-terminally deleted ( $\Delta$ 2-36) E1A 12S-

expression plasmid as indicated. These plasmids were constructed by subcloning FLAG-P/CAF and E1A cDNAs into pCX (34) and pcDNA1 (Invitrogen), respectively. All samples, in addition, contained 1  $\mu$ g of sorting plasmid (pCMV-IL2R) (31) and carrier plasmid (pCX) to normalize the total amount of DNA to 11  $\mu$ g. After  
5 transfection, cells were incubated in Dulbecco's modified Eagle's medium with 10% fetal bovine calf serum for 12 h, and subsequently labeled in medium containing 10  $\mu$ M bromo-deoxyuridine (BrdU) for 30 min. Subsequently, the transfected subpopulation was purified by magnetic affinity cell sorting and nuclei were analyzed by dual parameter flow cytometry as described (32).

10

The fraction of cells accumulating in S phase in control cultures was 23%, compared to 15% in P/CAF-transfected cells. This effect was reproducible in multiple independent experiments. In parallel experiments to verify the utility of this experimental protocol, plasmids encoding E2F-1, simian virus 40 small t, cyclin A or  
15 cyclin E increased the accumulation of cells in S phase, whereas plasmids encoding the cyclin-dependent kinase inhibitors p21 or p27 reduced the number of S phase cells.

On the basis of evidence that E1A and P/CAF compete for binding sites on p300, it seemed possible that cotransfection of P/CAF with E1A would oppose the  
20 mitogenic effect caused by E1A. As shown by the data herein, this is indeed the case. E1A alone has mitogenic activity in this experimental setting, while the E1A mutant lacking the p300 binding domain (E1A $\Delta$ N) has very weak activity. Comparable expression levels between wild type and mutant E1A in the transfected cells were revealed by immunoblotting analysis with anti-E1A. Intriguingly, when P/CAF was  
25 cotransfected with E1A, the mitogenic activity of E1A was significantly counteracted by P/CAF. These results show that P/CAF and E1A mediate antagonistic effects on cell cycle progression.

In the course of assessing P/CAF activity, it was also revealed that p300 is able  
30 to inhibit cell cycle progression under the same assay conditions. These findings suggest

that P/CAF and p300, perhaps by forming a complex, act in concert to suppress cell cycle progression.

#### **Histone acetyltransferase activity in P/CAF**

5        Acetylation of the N-terminal histone tails has been considered to play a crucial role in accessibility of transcription factors to nucleosomal templates (26-27). Recently, yGCN5 has been identified as a histone acetyltransferase (28). On the basis of this information, intrinsic histone acetyltransferase activity in P/CAF and hGCN5 was examined. As substrates, the core histones (histones H2A, H2B, H3 and H4) and the  
10    nucleosome core particles (146 base pairs of DNA wrapped around the octamer of core histones) were used.

Activity of hGCN5 and P/CAF that acetylates free histones or histones in the nucleosome core particle (35) was measured as described (36). Each reaction contained  
15    0.3 pmol of affinity purified FLAG-hGCN5 or FLAG-P/CAF, 4 pmol of the histone octamer or the nucleosome core particle and 10 pmol of [1-<sup>14</sup>C]acetyl-CoA. The histone octamer dissociated into dimers or tetramers under assay conditions. Acetylated histones were detected by autoradiography after separation by SDS-PAGE.

20        P/CAF and hGCN5 acetylated the core histones with almost the same efficiency. Both factors acetylated histones H3 and H4, but preferentially H3. In contrast, very weak or no acetylation by hGCN5 was detected in the nucleosome core particles. Remarkably, significant acetylation by P/CAF was observed in a nucleosomal context. Although all core histones are acetylated in the nucleus, P/CAF and hGCN5 did not  
25    acetylate histones H2A and H2B *in vitro*.

Direct function of P/CAF is likely to involve its intrinsic histone acetyltransferase activity. Although exact molecular mechanisms by which acetylation of core histones contribute to transcription remains undefined, acetylation of the histones is considered to  
30    play an important role in transcriptional regulation (26-27). The positively charged N-terminal tails of core histones are believed to affect nucleosome structure by interacting

with DNA at or near the nucleosome-spacer junction. Acetylation of the histone tails presumably destabilizes the nucleosome and facilitates access by regulatory factors. Likewise, there is a general correlation between the level of acetylation and transcriptional activity of nucleosomal domains. The findings of the present invention  
5 provide insights into the mechanisms of targeted histone acetylation.

Cellular factor p300/CBP binds to various sequence-specific factors that are involved in cell growth and/or differentiation, including CREB (3,4), c-Jun (9), Fos (11), c-Myb (12) and nuclear receptors (13). P/CAF could stimulate the activation function  
10 of these factors via promoter-specific histone acetylation. The present invention demonstrates that E1A appears to perturb normal cellular regulation by disrupting the connection between p300/CBP and its associated histone acetyltransferase.

## II. p300/CBP studies.

15

### **Purification of E1A associated histone acetyltransferase.**

FLAG-epitope tagged E1A (or  $\Delta$ E1A) was expressed in Sf9 cells (ATCC accession number CRL 1711) by infecting recombinant baculovirus (43). All purification steps were carried out at 4°C. Extract was prepared from infected cells by one cycle of  
20 freeze and thaw in buffer B (20 mM Tris-HCl, pH 8.0; 5 mM MgCl<sub>2</sub>; 10% glycerol; 1 mM PMSF; 10 mM  $\beta$ -mercaptoethanol; 0.1% Tween 20) containing 0.1 M KCl and the complete protease inhibitor cocktail (Boehringer Mannheim). To prepare E1A-immobilized beads, the extract was incubated with M2 anti-FLAG antibody agarose (Kodak-IBI) for four hours with rotating and  
25 subsequently washed with the same buffer three times. The resulting beads were incubated with HeLa (ATCC accession number CCL 2) nuclear extract for four to eight hours and thereafter washed with the same buffer six times. Finally, FLAG-E1A was eluted from the beads along with associated polypeptides by incubating with the same buffer containing 0.1 mg/ml FLAG peptide.

30

For further purification, eluted polypeptides were dialyzed in 0.05 M KCl-buffer B and subsequently loaded onto a SMART Mono Q column (Pharmacia) equilibrated with the same 0.05 M KCl-buffer B. After washing, the column was developed with a linear gradient of 0.05-1.0 M KCl in buffer B. Mono Q fractions were concentrated with a MICROCON spin-filter (Amicon) and consequently loaded onto a SMART Superdex 200 column (Pharmacia) equilibrated with 0.1 M KCl-buffer B.

#### Histone acetyltransferase assays

Filter binding assays were performed as described (80) with minor modifications. Samples were incubated at 30°C for 10-60 minutes in 30 ml of assay buffer containing 50 mM Tris-HCl, pH 8.0; 10% glycerol; 1 mM DTT; 1 mM PMSF; 10 mM sodium butyrate; 6 pmol of [<sup>3</sup>H]acetyl CoA (4.3 mCi/mmol, Amersham Life Science Inc.); and 33 mg/ml of calf thymus histones (Sigma Chemical Co.). In experiments where synthetic peptides were substituted for core histones, 50 pmol of each peptide were used. After incubation, the reaction mixture was spotted onto Whatman P-81 phosphocellulose filter paper and washed for 30 minutes with 0.2 M sodium carbonate buffer pH 9.2 at room temperature with 2-3 changes of the buffer. The dried filters were counted in a liquid scintillation counter.

PAGE analysis was done as above except that 90 pmol of [<sup>14</sup>C]acetyl CoA (55 mCi/mmol, Amersham Life Science Inc.) and 9 pmol of core histones or mononucleosomes were used. Core histones and mononucleosomes were prepared as described (35). For trypsin digestion, reaction mixtures were further incubated with various amounts of trypsin on ice for 30 minutes. The samples were analyzed on one dimensional SDS-PAGE gels or two dimensional gels, where the first dimension was an acid-urea-PAGE gel (44) and the second dimension was an SDS-PAGE gel.

#### Protein expression

For baculovirus expression, cDNAs corresponding to p300 portions of aa 1-670, aa 671-1194 and aa 1135-2414 were amplified by PCR (EXPAND High Fidelity PCR System; Boehringer Mannheim) as KpnI-NotI fragments. The resulting fragments were

subcloned into a baculovirus transfer vector having the FLAG-tag sequence (43). The recombinant viruses were isolated using the BACULOGOLD system (Pharmingen), according to the manufacturer's protocol and were infected into Sf9 cells (ATCC accession number CRL 1711) to express FLAG-p300. Recombinant proteins were  
5 affinity purified with M2 anti-FLAG antibody-immobilized agarose (Kodak-IBI) according to the manufacturer's protocol.

For bacterial expression, cDNAs encoding the p300 portions and the CBP portion (aa 1174-1850) were first subcloned into the baculovirus transfer vector having  
10 the FLAG-tag as described above. Thereafter, the XhoI and NotI fragments encoding FLAG-p300 or FLAG-CBP fusions were resubcloned into the *E. coli* expression vector pET-28c (Novagene) digested with SalI and NotI. Recombinant proteins were expressed in *E. coli* BL21(DE3) and affinity purified with M2-antibody agarose.

#### 15 **Histone acetyltransferases that associate with E1A**

Although the adenovirus E1A 12S protein (E1A) inhibits transcription in a variety of genes via direct binding to p300/CBP (45), E1A also stimulates transcription in some contexts (46). Thus, p300/CBP-bound E1A was tested to determine whether it might recruit histone acetyltransferases or deacetylases to regulate transcription. In  
20 addition, experiments were conducted as described below to determine if p300/CBP per se is a histone acetyltransferase.

Initially, recombinant FLAG-epitope tagged E1A was immobilized on anti-FLAG antibody beads. Immobilized E1A was incubated with a HeLa nuclear  
25 extract for affinity purification of E1A-associated polypeptides. FLAG-E1A was then eluted from the beads, along with E1A-associated polypeptides, by incubating with FLAG-peptide. Although E1A per se has no histone acetyltransferase activity, E1A recruited significant amounts of histone acetyltransferase activity from the nuclear extract. It is very unlikely that this activity is derived from P/CAF given that  
30 E1A and P/CAF cannot bind to p300/CBP simultaneously (43). Consistent with this, no P/CAF was detected in these fractions by immunoblotting.

The E1A N-terminus, a region that is not highly conserved among the various adenovirus serotypes, is involved in p300/CBP binding *in vivo*. Mutations in the N-terminal region lead to loss of the ability for p300/CBP binding without affecting RB binding (1,47). Thus, the requirement of the E1A N-terminal region for the recruitment of histone acetyltransferase activity was tested. In contrast to the wild type, the N-terminal deleted form of E1A ( $\Delta$ N-E1A) recruited only a background level of acetyltransferase activity. In agreement with previous reports (47), the  $\Delta$ N-E1A showed no ability to interact with p300/CBP, although it still retained the ability to interact with a variety of other polypeptides, including RB.

To define the relationship between p300/CBP and histone acetyltransferase activity, affinity purified E1A-binding polypeptides were separated by Mono Q ion-exchange column. Both p300/CBP and the acetyltransferase activity were coeluted at 140 mM KCl, while most of polypeptides were eluted at 260 mM KCl. The active fraction of Mono Q column (~140 mM KCl) was further separated by Superdex-200 gel filtration column. Both p300/CBP and the acetyltransferase activity coeluted after the void volume, indicating that p300/CBP is involved in the histone acetyltransferase activity.

#### **p300 is a histone acetyltransferase**

The data provided herein indicate that p300 per se, or a polypeptide(s) associated with p300, possesses histone acetyltransferase activity. To test the former possibility, the acetyltransferase activity of recombinant p300 was measured. p300 was divided into three fragments, each of which was expressed in and purified from Sf9 cells via a baculovirus expression vector. Histone acetyltransferase activity was readily detected in the C-terminal fragment containing amino acids 1135-2414, whereas no activity was found in the other fragments, demonstrating conclusively that p300 per se is a histone acetyltransferase.

**p300/CBP-histone acetyltransferase domain**

To map the histone acetyltransferase domain of p300, a series of deletions was prepared. Given the poor conservation of the glutamine-rich region (aa 1815-2414) in the *C. elegans* p300/CBP homolog (6), the p300 fragment encoding aa 1135-1810 was expressed in and purified from *E. coli*. Importantly, this candidate region of p300 (aa 1135-1810) showed significant histone acetyltransferase activity. For further mapping within this region, a series of N-terminal deletions was constructed. Deletion of 60 residues, resulting in a fragment containing aa 1195-1810, had no effect on the acetyltransferase activity, whereas the deletion of 185 residues, yielding a fragment comprising aa residues 1320-1810, completely eliminated the acetyltransferase activity.

Next, a series of C-terminal deletions was analyzed to determine the requirement of the P/CAF (or E1A) -binding domain. The p300 fragments lacking the E1A binding domain (aa 1195-1760, 1195-1706 and 1195-1673) still retained the acetyltransferase activity, whereas the further truncated mutant (aa 1195-1652) completely lost the acetyltransferase activity. Consistent with these results, the internal deletion of residues 1418-1720 showed no acetyltransferase activity. These data demonstrate that the histone acetyltransferase domain is located between the bromodomain and the -E1A-binding domain. Given that the histone acetyltransferase domain is highly conserved between p300 and CBP (91% similarity), the corresponding region of CBP, aa residues 1174-1850, was expressed to confirm the acetyltransferase activity. As expected, comparable activity was detected, indicating that both p300 and CBP are histone acetyltransferases.

Among various acetyltransferases including histone acetyltransferases GCN5 and P/CAF, putative acetyl-CoA binding sites are conserved (48). However, multiple alignment analysis (49) showed that the p300/CBP histone acetyltransferase domain does not belong to this group. Moreover, comparison of the p300/CBP histone acetyltransferase domain with peptide sequence databases (23) showed no sequence similarity to any other proteins. Accordingly, this invention shows that p300/CBP represents a novel class of acetyltransferases in that it does not have the conserved motif found among previously described acetyltransferases (48).

**p300 acetylates all core histones in mononucleosomes**

Substrate specificity for acetylation by p300 was also examined. As substrates, histone octamers and mononucleosomes (146 base pairs of DNA wrapped around the octamer of core histones) were used. Given that the histone octamer dissociates into  
5 dimers or tetramers under physiological conditions, the histone octamer is referred to here as core histones. When core histones were used, p300 acetylated all four proteins, but preferentially H3 and H4. More importantly, in a nucleosomal context, p300 acetylated all four core histones nearly stoichiometrically. In contrast, p300 acetylated neither BSA nor lysozyme.

10

Hyperacetylated histones are believed to be linked with transcriptionally active chromatin (26,27,50,51). Hyperacetylated forms are found in histones H4, H3 and H2B, which have multiple acetylation sites *in vivo*. Thus, the level of acetylation by p300 was also tested.

15

Mononucleosomes treated with p300 were analyzed by two-dimensional gel electrophoresis. A Coomassie blue-stained gel and the corresponding autoradiogram showed that a significant amount of histones, especially H4, were hyperacetylated. Importantly, acetylation levels by p300 were very close to those of hyperacetylated  
20 histones prepared from HeLa nuclei treated with sodium butyrate, a histone deacetylase inhibitor. In contrast, no acetylated forms were detected in the reaction without p300. These results indicate that p300 acetylates histones in mononucleosomes to the hyperacetylated state by targeting multiple lysine residues.

25

**p300 acetylates the four lysines in the histone H4 N-terminal tail *in vitro* which are acetylated *in vivo***

Lysines at positions 5, 8, 12 and 16 of histone H4 are acetylated *in vivo* (51). Recent studies with yeast histone acetyltransferases demonstrate the  
30 position-specific acetylation by distinct acetyltransferases, i.e., while cytoplasmic acetyltransferases for histone deposition and chromatin assembly modify

positions 5 and 12, GCN5 modifies positions 8 and 16 (52). Accordingly, the positions of acetylation by p300 were also determined. A series of synthetic peptides containing acetylated lysines at various positions was used to determine the acetylation site-specificity of p300. Consistent with the two-dimensional gel electrophoresis analysis, the experiments with peptide substrates showed that p300 acetylates all four lysines in the histone H4 that are acetylated *in vivo*. These results are consistent with the view that deposition-related diacetylated histones are deacetylated during maturation of chromatin (53).

#### 10 p300 preferentially acetylates the N-terminal histone tail

Histone acetyltransferases modify specific lysine residues in the N-terminal tail of core histones but not the C-terminal globular domain *in vivo* (26,27,50,51). Structural models of nucleosomes (54,55,56) suggest that most of the lysine residues in the C-terminal globular domain are buried. Therefore, experiments were conducted to examine whether restricted acetylation of the N-terminal tail resulted from the substrate specificity of the enzyme or inaccessibility of the enzyme to the core domain in nucleosomes. The globular domains of all core histones contain a long helix flanked on either side by a loop segment and short helix, termed the "histone fold" (54,55,56). The histone fold is involved in formation of the stable H2A-H2B and H3-H4 hetero-dimers, consisting of extensive hydrophobic contacts between the paired molecules. Therefore, it is likely that a histone monomer cannot fold properly, thereby increasing access of the histone acetyltransferase to the core domain. Based on these considerations, experiments were conducted to determine whether p300 acetylates free histone H4 in a N-terminal-specific manner.

25

Histone H4 was acetylated with p300 and subsequently the histone tail was removed by partial digestion with trypsin. The distributions of radioactivity between intact and core histones were compared. While the globular core histone domain was predominant at the higher trypsin concentrations, radioactivity was detected mostly in the intact histone. These data demonstrate that p300 preferentially acetylates the N-terminal tail of histone H4.

30

### 5 III. P/CAF interaction with MyoD

#### Tissue culture and transfection experiments

C<sub>2</sub>C<sub>12</sub> mouse cells (ATCC accession number CRL 1772) were grown in Dulbecco's modified Eagle medium (DMEM) supplemented with 20% fetal bovine serum (FBS) until they reached confluence. Differentiation was induced by switching  
10 medium to differentiation medium (DM), consisting of DMEM containing 2% horse serum. C<sub>3</sub>H/10T1/2 fibroblasts (ATCC accession number CCL 226) were grown in DMEM supplemented with 10% FBS. Cells were transfected by the calcium phosphate precipitation method. Total amounts of transfected DNA were equalized by empty  
15 vector DNA. After 12 h incubation in medium containing the precipitated DNA, the cells were washed and incubated in fresh DMEM containing 10% FBS for an additional 24 h. Afterwards, differentiation was induced by incubating in DM for 36 to 72 h. Chloramphenicol acetyltransferase (CAT) assays were performed as previously described (64,69). The quantities of cell extracts used for CAT assays were normalized  
20 to  $\beta$ -galactosidase activity by cotransfection of 1 mg of the  $\beta$ -galactosidase expression vector, pON260.

Expression vectors used for transfection experiments are as follows:

pCX-P/CAF for P/CAF (43); pCMV-bp300 for p300 (65), pCMV-p300 (1869-2414)  
25 (64) and pCMV-p300 (1514-1922) (60) for p300 wild type and mutants; pE1A12S, pE1A12S R2G, pE1A12S D2-36 and pE1A12S D121-130 for E1A wild type and mutants (66,67,68); and pEMSV-MyoD for MyoD (64).

The antisense P/CAF RNA expression vector, pcDNA3 P/CAF-AS, was created  
30 as follows. The 2.5 Kb EcoRI-KpnI fragment containing the entire P/CAF open reading frame was isolated from pCX-P/CAF (43). This fragment was subcloned into the

EcoRI-KpnI sites of plasmid pcDNA3 (Invitrogen) so that the antisense P/CAF RNA is driven under the CMV promoter. Reporter genes employed were 4RE-CAT and MCK-CAT (69). 4RE-CAT is driven by a synthetic promoter containing 4 copies of the E-box, whereas MCK-CAT is driven by the native MCK promoter (nucleotides -1256 to +7).

### Microinjection and immunofluorescence

Cells were grown on small glass slides, subdivided into numbered squares of 2 mm x 2 mm and microinjected with purified and concentrated antibodies, as previously described (70). For immunofluorescence, cells were fixed in either 2% paraformaldehyde or 1:2 methanol/acetone solution, preincubated with 5% BSA/PBS and incubated with the primary antibodies for 30 min at 37° C. Subsequently, antibody was visualized by incubating with either rhodamine- or fluorescein-conjugated secondary antibody for 30 min at 37° C. Injected antibodies were stained with a rhodamine-conjugated secondary antibody and nuclei were counter-stained by DAPI as previously described (69).

Antibodies employed are as follows; rabbit polyclonal affinity purified anti-P/CAF antibody (43), rabbit polyclonal anti-p300/CBP antiserum (71), mouse monoclonal anti-MyoD antibody (clone 5.8A, kindly provided by P. Houghton), goat polyclonal anti-c-Jun affinity purified antibody (Santa Cruz) and rabbit pre-immune serum.

25

### Immunoprecipitation and DNA affinity purification

Cells were resuspended in lysis buffer (20 mM NaPO<sub>4</sub>, 150 mM NaCl, 5mM MgCl<sub>2</sub>, 0.1% NP40, 1 mM DTT, 10 mM sodium fluoride, 0.1 mM sodium vanadate, 1 mM phenylmethylsulfonyl-fluoride and 10 mg/ml each of leupeptin, aprotinin and pepstatin). After 30 min incubation on ice, samples were centrifuged at 12,000 x g for 30 min and supernatants were used as cell extracts. Extracts were pre-cleared by

incubating with rabbit pre-immune serum and protein A/G Plus-Agarose (Santa Cruz) for 2 h at 4 C. For immunoprecipitation, the supernatants were incubated with the respective antibodies for 3 h at 4 C. Protein A/G Plus-Agarose was added, and incubation continued for 3 h. The matrix was washed with lysis buffer, then boiled in 2 X SDS sample buffer. Immunoblotting was performed by using the ECL chemiluminescent detection kit (Amersham) according to the manufacturer's protocol.

Affinity purification of E-box-bound complexes was done as previously described (69). Briefly, 100 ng of the biotinylated double stranded DNA containing the E-box were immobilized on streptavidin-conjugated magnetic beads and incubated with 500 mg of cell extracts in the presence of poly dI-dC. After extensive washing, bound proteins were eluted with SDS sample buffer and analyzed by immunoblotting.

#### **In vitro protein-protein interaction assays**

The CBP-B fragment and its deletion derivatives were expressed as GST-fusions described previously (43). MyoD and E1A (43) were expressed as FLAG-fusion proteins in Sf9 cells via a baculovirus expression system and affinity-purified on M2 anti-FLAG antibody-agarose (Kodak-IBI). Crude E. coli extracts containing GST-fusions were incubated with various amounts of MyoD and/or E1A in 50 ml of buffer B (20 mM Tris-HCl, pH 8.0, 0.1 M KCl, 5 mM MgCl<sub>2</sub>, 10% glycerol, and 0.1% Nonidet P-40) on ice for 10 min. GST-precipitation was performed as described (43). MyoD and E1A were detected by immunoblotting with anti-FLAG M2 antibody. For the interaction between P/CAF and MyoD, 1.5 pmol of FLAG-P/CAF and 15 pmol of FLAG-MyoD were incubated in 50 ml of buffer B on ice for 10 min. The mixture was further incubated with 2 mg of anti-P/CAF (43) or anti-hADA2 antibody for 60 min. The immunocomplexes were precipitated by incubation with 10 ml of protein A-Trisacryl (Pierce) and rotated for 1-4 hr at 4°C. The matrix was washed 4 times with 200 ml of buffer B and boiled in 10 ml of 2 X SDS sample buffer. The proteins were resolved on a 4%-20% gradient SDS-PAGE and subjected to immunoblotting with the anti-FLAG M2 antibody. The blot was developed with the SUPERSIGNAL chemiluminescent substrates (Pierce).

### **P/CAF coactivates muscle-specific transcription**

P/CAF and MyoD were co-transfected into mouse C3H10T1/2 fibroblasts, and MyoD-mediated transcription was determined from reporter activity driven by the artificial (4RE) and the naturally-occurring muscle creatine kinase (MCK) promoters. Overexpression of P/CAF stimulated MyoD-dependent transcription several folds in both promoters. Similar results were obtained for the myoD activated myogenin promoter. Transcriptional activation was further stimulated by co-transfecting with MyoD, P/CAF and p300 expression vectors, suggesting that P/CAF may function by forming a complex with p300/CBP. Consistent with the lack of DNA binding capacity in P/CAF, overexpression of P/CAF alone did not increase the basal transcriptional activity of either enhancer. To test whether P/CAF and p300/CBP function in the same pathway, two dominant negative forms of p300 were employed which specifically inhibit p300/CBP-mediated transcription (60,64). The p300 segment spanning residues 1514-1922 inhibits the MyoD-dependent activation via direct interaction with MyoD (60), whereas the p300 segment spanning residues 1869-2414 inhibit it without direct interaction (64). Both dominant negative mutants inhibited MyoD-coactivation by P/CAF, suggesting that P/CAF and p300/CBP function in the same pathway.

For further elucidation of the activation mechanism by P/CAF, the effect of E1A, which inhibits MyoD-dependent transcription and differentiation (66,72,73) via direct interaction with p300/CBP (65,78), was tested. Expression of E1A in C3H10T1/2 fibroblasts inhibited stimulation of MyoD-directed transcription by P/CAF overexpression. E1A mutants lacking p300/CBP-binding activity, E1A D2-36 and E1A R2G (67,79), had almost no effect. On the other hand, an E1A mutant retaining p300/CBP-binding activity, E1A D121-130, behaved like the wild type. Since E1A associates with p300/CBP, but not with P/CAF, these results suggest that P/CAF functions in MyoD-directed transcription via interaction with p300/CBP.

To address the role of P/CAF as a myogenic coactivator in a more relevant environment, P/CAF was overexpressed in proliferating C2C12 myoblasts which express

endogenous myogenic bHLH factors. As observed in fibroblasts, overexpression of P/CAF stimulated muscle specific transcription. Concomitant expression of exogenous p300 increased P/CAF-mediated coactivation. The repression exerted by wild type E1A, but not mutant E1A D2-36, on P/CAF coactivation of MyoD was also observed in  
5 muscle cells.

Similar experiments were performed with myogenic cell lines that were stably transformed with wild type or mutant E1A-expressing vectors (66). Coactivation by P/CAF was inhibited by wild type E1A or the E1A mutant that retains  
10 p300/CBP-binding activity (E1A $\Delta$ 121-130). In contrast, E1A mutants that lack p300/CBP-binding (E1A  $\Delta$ 2-36 and E1A R2G) allowed transcriptional coactivation by P/CAF. Taken together, these experiments show that P/CAF coactivates MyoD-directed transcription via interaction with p300/CBP.

#### 15 **P/CAF stimulates myogenic differentiation**

Given that P/CAF potentiates MyoD-directed transcription, the ability of P/CAF to assist MyoD in promoting myogenic differentiation was investigated. To this aim, C3H10T1/2 fibroblasts were transiently transfected with P/CAF and MyoD expression vectors. An expression vector for the green fluorescent protein (GFP) was  
20 co-transfected to identify transfected cells. After incubation in differentiation medium, the myogenic conversion of transfected cells was determined by simultaneous expression of the GFP and the differentiation-specific marker myosin heavy chain (MHC). Forced expression of MyoD in fibroblasts caused muscle differentiation in 12% of the transfected fibroblasts. This myogenic conversion was 20% by co-expressing MyoD and  
25 P/CAF. As observed in transcription experiments, stimulation of differentiation by P/CAF was counteracted by co-transfection with the p300 dominant negative mutant, p300 (1869-2414). Consistent with a general role for coactivators, overexpression of P/CAF alone was unable to differentiate fibroblasts.

30 Similar experiments were done using proliferating C2C12 myoblasts in which the differentiation program is already committed. Most of the myoblasts differentiated into

myotubes by overexpressing P/CAF, whereas only a modest effect was observed by overexpressing p300. In contrast, differentiation was inhibited slightly by overexpressing c-Jun. This inhibitory effect presumably was caused by titration of p300/CBP, which associates directly with c-Jun (74). A similar inhibition was observed in the p300 dominant negative mutant. Consistent with the transcriptional effect, E1A almost completely inhibited differentiation. The E1A mutant RG2, lacking p300/CBP-binding capability but retaining the retinoblastoma protein (Rb)-binding capability, only partially inhibited differentiation, although this same mutant inhibited transcription as severely as the wild type. Taken together, these data show that P/CAF stimulates muscle differentiation by coactivating MyoD function via p300/CBP association.

#### **P/CAF is essential for myogenic transcription and differentiation**

To test the necessity of P/CAF for myogenic transcription, experiments were conducted whereby P/CAF synthesis was inhibited by expressing antisense P/CAF RNA. A vector from which the P/CAF mRNA is transcribed in the antisense orientation (P/CAF-AS) was transfected with P/CAF and MyoD expression vectors into fibroblasts and MyoD-dependent transcription was examined. Cotransfection of the antisense expression vector strongly inhibited MyoD-dependent transcription below the level of induction elucidated by MyoD alone, demonstrating that expression of P/CAF antisense RNA inhibits not only the coactivation exerted by exogenous P/CAF but also that of endogenous P/CAF. These results indicate that P/CAF is essential for MyoD-dependent transcription.

Studies were also carried out to determine whether expression of P/CAF antisense RNA inhibits myogenic differentiation. C3H10T1/2 fibroblasts were transiently transfected with various expression vectors with or without the P/CAF antisense RNA expression vector. Expression of P/CAF antisense RNA reduced MyoD-mediated myogenic conversion of fibroblasts. Expression of P/CAF antisense RNA also counteracted the stimulatory effect of both P/CAF and p300 on myogenic differentiation. These data support the view that P/CAF and p300/CBP coactivate

MyoD-dependent transcription in the same pathway. More drastic inhibition was observed in C2C12 myoblasts in similar experiments. Therefore, it can be concluded that P/CAF is essential for transcription of muscle specific genes and hence differentiation into myotubes.

5

To further confirm the essential role of P/CAF for myogenic differentiation, we blockage experiments by antibody microinjection were performed. Antibodies were injected into the cytoplasm of proliferating C2C12 myoblasts to prevent the nuclear transport of newly synthesized target proteins. After incubating in the differentiation medium, the degree of differentiation was determined. Microinjection of an anti-P/CAF antibody almost completely inhibited differentiation. Similar results were obtained by microinjecting anti-p300/CBP antibodies. Although microinjection of either anti-p300/CBP or P/CAF antibody was sufficient to inhibit differentiation, an even greater inhibition was observed by coinjecting both of them. Microinjection of anti-P/CAF or anti-p300/CBP antibody did not interfere with induction of p53 by DNA damaging agents, showing specificity of the inhibition by the antibodies. In contrast to anti-P/CAF or anti-p300/CBP antibodies, the injection of anti-MyoD antibody only partially inhibited differentiation, supporting the view of functional redundancy between MyoD and Myf-5 (75,76). Injection of anti-c-Jun antibody or control antibody did not interfere with muscle differentiation.

20

Similar experiments were performed with C3H10T1/2 fibroblasts stably expressing MyoD. In these cells, either anti-p300/CBP or anti-P/CAF antibody completely inhibited muscle differentiation. In contrast to myoblasts, anti-MyoD antibody completely blocked differentiation in the fibroblasts expressing MyoD. Anti-c-Jun and control antibodies did not interfere with differentiation. Taken together, these results demonstrate that P/CAF and p300/CBP are indispensable for activation of the myogenic program.

25

30

**p300/CBP, P/CAF and MyoD form a multimeric complex *in vivo***

The data described above indicate that P/CAF stimulates MyoD-directed transcription via association with p300/CBP. Thus, experiments were conducted to investigate whether P/CAF, p300/CBP and MyoD could associate in a complex.

- 5 First, cellular extracts derived from C2C12 myotubes were subjected to immunoprecipitation. Both anti-MyoD and anti-p300/CBP antibodies co-precipitated P/CAF. In a complementary experiment, both anti-p300/CBP and anti-P/CAF antibodies also co-precipitated MyoD, suggesting that these factors form a multimeric protein complex in myotubes.

10

- Next, attempts were made to detect this complex on the E-box, the DNA binding site for MyoD. Immobilized DNA containing an E-box sequence was incubated with myotube extracts. After extensive washing, P/CAF, p300/CBP and MyoD were analyzed by immunoblotting. P/CAF, p300/CBP and MyoD were all affinity purified on  
15 the immobilized DNA, whereas they were not purified on the control DNA lacking the E-box. Given that P/CAF and p300/CBP per se cannot bind to DNA, these observations indicate that P/CAF and p300/CBP are recruited through MyoD at the E-box sites to form a multi-protein complex.

**20 Complex formation is inhibited by viral transforming factors**

- Since the oncoviral proteins E1A and large T antigen inhibit myogenic transcription and differentiation, the effect of these factors on the formation of complexes on the E-box was tested. Importantly, very small amounts of P/CAF and p300/CBP were co-purified on the E-box from myocyte extracts which stably express  
25 E1A or large T antigen, although MyoD was detected under these conditions. The lower recovery of MyoD from E1A-expressing muscle cells could reflect the low level of MyoD in the extracts (66). These results indicate that E1A and large T antigen dissociate P/CAF and p300/CBP from MyoD without altering MyoD binding to DNA.

- 30 Consistent with the previous observations that transiently expressed E1A prevents interaction between P/CAF and p300/CBP *in vivo* (43), the association

between p300/CBP and P/CAF was abolished in myoblasts stably transformed by wild type E1A but not in those clones transformed with the E1A mutant R2G unable to bind p300/CBP. Similarly, the interaction between p300/CBP and P/CAF was abolished by large T antigen but not by the mutant protein that localizes into the cytoplasm (77).

5

#### **Interaction between MyoD, P/CAF and CBP *in vitro***

Previous interaction experiments *in vitro* indicate that the CBP region spanning residues 1801 to 1850 is crucial for interaction with both P/CAF and E1A (43). While most sequence-specific factors bind to CBP sites distinct from the P/CAF/E1A binding sites, MyoD interacts with an overlapping CBP fragment called the CH3 region (60,64,65). To understand how P/CAF, p300/CBP and MyoD associate, the CBP sites important for MyoD binding were mapped more precisely. Consistent with previous reports (60,64,65), the CBP fragment spanning residues 1801-2000 (fragment B) bound MyoD. Moreover, deletion of residues 1801 to 1850 within fragment B completely abolished interaction with MyoD, which is similar to the results obtained with P/CAF and E1A. Importantly, an internal deletion of residues 1850-1878 abolished the MyoD interaction with CBP, while it did not affect binding of E1A or P/CAF (43). These results suggest that MyoD and P/CAF bind to distinct sites of p300/CBP, albeit the binding sites may overlap. Moreover, a direct interaction was observed between MyoD and P/CAF, which may contribute to stabilization of the multimeric complex.

15  
20

These data show that E1A prevents not only p300/CBP-interaction with P/CAF but also that with MyoD *in vivo*. To obtain evidence that this inhibition is due to the direct action by E1A, competition experiments were performed *in vitro*. Importantly, the interaction between CBP and MyoD was strongly inhibited by addition of E1A, implicating that E1A inhibits myogenic transcription by disrupting multiple interactions.

25

Although the present process has been described with reference to specific details of certain embodiments thereof, it is not intended that such details should be

30

regarded as limitations upon the scope of the invention except as and to the extent that they are included in the accompanying claims.

Throughout this application various publications are referenced by numbers within parentheses. Full citations for these publications are as follows. The disclosures of these publications in their entireties are hereby incorporated by reference into this application in order to more fully describe the state of the art to which this invention pertains.

### References

1. Moran, E. *Curr. Opin. Genet. Dev.* **3**, 63-70 (1993).
2. Dyson, N. & Harlow, E. *Cancer Surv.* **12**, 161-195 (1992).
3. Chrivia, J. C., *et al.* *Nature* **365**, 855-859 (1993).
4. Kwok, R. P., *et al.* *Nature* **370**, 223-226 (1994).
5. Eckner, R., *et al.* *Genes Dev.* **8**, 869-884 (1994).
6. Arany, Z., Sellers, W. R., Livingston, D. M. & Eckner, R. *Cell* **77**, 799-800 (1994).
7. Arany, Z., Newsome, D., Oldread, E., Livingston, D. M. & Eckner, R. *Nature* **374**, 81-84 (1995).
8. Lundblad, J. R., Kwok, R. P., Lurance, M. E., Harter, M. L. & Goodman, R. H. *Nature* **374**, 85-88 (1995).
9. Arias, J., *et al.* *Nature* **370**, 226-229 (1994).
10. Lee, J. S., *et al.* *Genes Dev.* **9**, 1188-1198 (1995).
11. Banister, A. J. & T., K. *EMBO J.* **14**, 4758-4762 (1995).
12. Dai, P. *et al.* *Genes Dev.* **10**, 528-540 (1996).
13. Kamei, Y. *et al.* *Cell* **85**, 403-414 (1996).
14. Rochette-Egly, C., Fromental, C., and Chambon, P. *Genes Dev.* **5**, 1200-1211 (1990).
15. Struhl, K. *Cell* **50**, 841-846 (1987).

16. Engelberg, D., Klein, C., Martinetto, H., Struhl, K. & Karin, M. *Cell* 77, 381-390 (1994).
17. Georgakopoulos, T. & Thireos, G. *EMBO J.* 11, 4145-4152 (1992).
18. Berger, S. L., *et al.* *Cell* 70, 251-265 (1992).
19. Horiuchi, J., Silverman, N., Marcus, G. A. & Guarente, L. *Mol. Cell. Biol.* 15, 1203-1209 (1995).
20. Berger, S. L., Cress, W. D., Cress, A., Triezenberg, S. J. & Guarente, L. *Cell* 61, 1199-1208 (1990).
21. Marcus, G. A., Silverman, N., Berger, S. L., Horiuchi, J. & Guarente, L. *EMBO J.* 13, 4807-4815 (1994).
22. Georgakopoulos, T., Gounalaki, N. & Thireos, G. *Mol. Gen. Genet.* 246, 723-728 (1995).
23. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. *J. Mol. Biol.* 215, 403-410 (1990).
24. Adams, M. D., Soares, M. B., Kerlavage, A. R., Fields, C. & Venter, J. C. *Nature Genet.* 4, 373-380 (1993).
25. Candau, R., *et al.* *Mol. Cell. Biol.* 16, 593-602 (1996).
26. Turner, B. M. *Cell* 75, 5-8 (1993).
27. Wolffe, A. P. *Trends Biochem. Sci.* 19, 240-244 (1994).
28. Brownell, J. E., *et al.* *Cell* 84, 843-851 (1996).
29. Haynes, S. R., *et al.* *Nucl. Acids Res.* 20, 2063 (1992).
30. Hopp, T. P., *et al.* *Biotechnology* 6, 1205-1210 (1988).
31. Giordano, T., Howard, T. H., Coleman, J., Sakamoto, K. & Howard, B. H. *Exp. Cell Res.* 192, 193-197 (1991).
32. Ogryzko, V. V., Hirai, T. H., Shih, C. E. & Howard, B. H. *J. Virol.* 68, 3724-3732 (1994).
33. Harlow, E. D. & Lane, D. *Antibodies: A laboratory manual* (Cold Spring Harbor Laboratory Press, New York, 1988).
34. Minucci, S., *et al.* *Mol. Cell. Biol.* 14, 360-372 (1994).
35. Hayes, J. J., Pruss, D. & Wolffe, A. P. *Proc. Natl. Acad. Sci. U. S. A.* 91, 7817-7821 (1994).

36. Brownell, J. E. & Allis, C. D. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 6364-6368 (1995).
37. Dayhoff et al. *Atlas of Protein Sequence and Structure 1978*, Nat'l Biomed. Res. Found., Washington, D.C.
38. Pei et al., *J. Biol. Chem.* **266**:16363-16369 (1991).
39. Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989.
40. Cunningham, et al. *Science* **243**:1330-1336 (1989).
41. Ferretti, et al. *Proc. Nat. Acad. Sci.* **82**:599-603 (1986).
42. Brake et al. *PNAS* **82**:4642-4646 (1984).
43. Yang, X.J., et al. *Nature* **382**:319-324 (1996).
44. Panyim, S. and Chalkley, R. *Arch. Biochem. Biophys.* **130**:337-346 (1969).
45. Song, C.Z., et al. *J. Biol. Chem.* **264**:24524-24530 (1989).
46. Kitabayashi, I., et al. *EMBO J.* **14**:3496-3509 (1995).
47. Wang, H.G., et al. *J. Virol.* **67**:4804-4813 (1993).
48. Lu, L., et al. RGFGIGS is an amino acid sequence required for acetyl coenzyme A binding and activity of human spermidine/spermine N1 acetyltransferase. *271*:18920-18924 (1996).
49. Lawrence, C.E., et al. *Science* **262**:208-214 (1993).
50. Wolffe, A.P. and Pruss, D. *Cell* **86**:817-819 (1996).
51. Brownell, J.E. and Allis, C.D. *Curr. Opin. Genet. Dev.* **6**:176-184 (1996).
52. Kuo, M.-H., et al. *Nature* **383**:269-272 (1996).
53. Shimamura, A. and Worcel, A. *J. Biol. Chem.* **264**:14524-14530 (1989).
54. Arents, G. and Moudrianakis, E.N. *PNAS* **90**:10489-10493 (1993).
55. Ramakrishnan, V. *Curr. Opin. Struct. Biol.* **4**:44-50 (1994).
56. Pruss, D., et al. *Bioassays* **17**:161-170 (1995).
57. Chakravarti et al. *Nature* **383**:99-103 (1996).
58. Oelgeschlager et al. *EMBO J.* **15**:2771-2780 (1996).
59. Janknecht and Nordheim *Oncogene* **12**:1961-1969 (1996).
60. Yuan, W., Condorelli, G., Caruso, M., Felsani, A., and Giordano, A. Human p300 protein is a coactivator for the transcription factor MyoD. *J.*

- Biol. Chem.* 271:9009-9013 (1996).
61. Onate et al. *Science* 270:1354-1357 (1995).
  62. A. Melton, Ed., *Antisense RNA and DNA*, D. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1988).
  63. Puri, P. L., Avantaggiati, M. L., Balsano, C., Sang, N., Graessmann, A., Giordano, A., and Levrero, M. p300 is required for MyoD-dependent cell cycle arrest and muscle-specific gene transcription. *EMBO J.* 16:369-383 (1997).
  64. Sartorelli, V., Huang, J., Hamamori, Y., and Kedes, L. Molecular mechanisms of myogenic coactivation by p300: direct interaction with the activation domain of MyoD and with the MADS box of MEF2C. *Mol. Cell. Biol.* 17:1010-1026 (1997).
  65. Eckner, R., Yao, T. P., Oldread, E., and Livingston, D. M. Interaction and functional collaboration of p300/CBP and bHLH factors in muscle and B-cell differentiation. *Genes Dev.* 10:2478-2490 (1996b).
  66. Caruso, M., Martelli, F., Giordano, A., and Felsani, A. Regulation of MyoD gene transcription and protein function by the transforming domains of the adenovirus E1A oncoprotein. *Oncogene* 8:267-278 (1993).
  67. Wang, H. G., Rikitake, Y., Carter, M. C., Yaciuk, P., Abraham, S. E., Zerler, B., and Moran, E. Identification of specific adenovirus E1A N-terminal residues critical to the binding of cellular proteins and to the control of cell growth. *J. Virol.* 67:476-488 (1993a).
  68. Wang, H. G., Yaciuk, P., Ricciardi, R. P., Green, M., Yokoyama, K., and Moran, E. The E1A products of oncogenic adenovirus serotype 12 include amino-terminally modified forms able to bind the retinoblastoma protein but not p300. *J. Virol.* 67:4804-4013 (1993b).
  69. Puri, P. L., Avantaggiati, M. L., Balsano, C., Sang, N., Graessmann, A., Giordano, A., and Levrero, M. p300 is required for MyoD-dependent cell cycle arrest and muscle-specific gene transcription. *EMBO J.* 16:369-383 (1997).
  70. Graessmann, M. and Graessmann, A. Microinjection of tissue culture

- cells. *Methods Enzymol.* 101:482-492 (1983).
71. Avantaggiati, M. L., Carbone, M., Graessmann, A., Nakatani, Y., Howard, B., and Levine, A. S. The SV40 large T antigen and adenovirus E1a oncoproteins interact with distinct isoforms of the transcriptional co-activator, p300. *EMBO J.* 15:2236-2248 (1996).
  72. Webster, K. A., Muscat, G. E., and Kedes, L. Adenovirus E1A products suppress myogenic differentiation and inhibit transcription from muscle-specific promoters. *Nature* 332:553-557 (1988).
  73. Mymryk, J. S., Lee, R. W., and Bayley, S. T. Ability of adenovirus 5 E1A proteins to suppress differentiation of BC3H1 myoblasts correlates with their binding to a 300 kDa cellular protein. *Mol. Biol. Cell.* 3:1107-1115 (1992).
  74. Arias, J., Alberts, A. S., Brindle, P., Claret, F. X., Smeal, T., Karin, M., Feramisco, J., and Montminy, M. Activation of cAMP and mitogen responsive genes relies on a common nuclear factor. *Nature* 370:226-229 (1994).
  75. Rudnicki, M. A., Braun, T., Hinuma, S., and Jaenisch, R. Inactivation of MyoD in mice leads to up-regulation of the myogenic HLH gene Myf-5 and results in apparently normal muscle development. *Cell* 71, 383-390 (1992).
  76. Rudnicki, M. A., Schnegelsberg, P. N., Stead, R. H., Braun, T., Arnold, H. H., and Jaenisch, R. MyoD or Myf-5 is required for the formation of skeletal muscle. *Cell* 75:1351-1359 (1993).
  77. Tedesco, D., Caruso, M., Fischer, F. L., and Vesco, C. The inhibition of cultured myoblast differentiation by the simian virus 40 large T antigen occurs after myogenin expression and Rb up-regulation and is not exerted by transformation-competent cytoplasmic mutants. *J. Virol.* 69:6947-6957 (1995).
  78. Moran, E. DNA tumor virus transforming proteins and the cell cycle. *Curr. Opin. Genet. Dev.* 3:63-70 (1993).
  79. Giordano et al. *Exp. Cell. Res.* 192:193-197 (1991).
  80. Brownell and Allis *PNAS* 92:6364-6368 (1995).

## SEQUENCE LISTING

## (1) GENERAL INFORMATION.

(i) APPLICANT: The United States of America, as represented by the Secretary, Department of Health and Human Services, c/o National Institutes of Health, Office of Technology Transfer, 6011 Executive Boulevard, Suite 325, Rockville, Maryland 20842

(ii) TITLE OF THE INVENTION: METHODS AND COMPOSITIONS FOR p300/CBP-ASSOCIATED TRANSCRIPTIONAL CO-FACTOR P/CAF

(iii) NUMBER OF SEQUENCES: 18

## (iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: NEEDLE & ROSENBERG, P.C.  
(B) STREET: Suite 1200, 127 Peachtree Street, NE  
(C) CITY: Atlanta  
(D) STATE: GA  
(E) COUNTRY: USA  
(F) ZIP: 30303

## (v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Diskette  
(B) COMPUTER: IBM Compatible  
(C) OPERATING SYSTEM: DOS  
(D) SOFTWARE: FastSEQ for Windows Version 2.0

## (vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:  
(B) FILING DATE: 23-JUL-1997  
(C) CLASSIFICATION:

## (vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: Corresponding U.S. Serial No. 60/022,273  
(B) FILING DATE: 23-July-1996

## (viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Miller, Mary L  
(B) REGISTRATION NUMBER: 39,303  
(C) REFERENCE/DOCKET NUMBER: 14014.0238/P

## (ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 404/688-0770  
(B) TELEFAX: 404/688-9880  
(C) TELEX:

## (2) INFORMATION FOR SEQ ID NO:1:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 832 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Met	Ser	Glu	Ala	Gly	Gly	Ala	Gly	Pro	Gly	Gly	Cys	Gly	Ala	Gly	Ala		
1				5				10					15				
Gly	Ala	Gly	Ala	Gly	Pro	Gly	Ala	Leu	Pro	Pro	Gln	Pro	Ala	Ala	Leu		
			20					25					30				
Pro	Pro	Ala	Pro	Pro	Gln	Gly	Ser	Pro	Cys	Ala	Ala	Ala	Ala	Gly	Gly		
			35				40						45				
Ser	Gly	Ala	Cys	Gly	Pro	Ala	Thr	Ala	Val	Ala	Ala	Ala	Gly	Thr	Ala		
			50				55					60					
Glu	Gly	Pro	Gly	Gly	Gly	Gly	Ser	Ala	Arg	Ile	Ala	Val	Lys	Lys	Ala		
65						70				75					80		
Gln	Leu	Arg	Ser	Ala	Pro	Arg	Ala	Lys	Lys	Leu	Glu	Lys	Leu	Gly	Val		
				85					90					95			
Tyr	Ser	Ala	Cys	Lys	Ala	Glu	Glu	Ser	Cys	Lys	Cys	Asn	Gly	Trp	Lys		
			100					105					110				
Asn	Pro	Asn	Pro	Ser	Pro	Thr	Pro	Pro	Arg	Ala	Asp	Leu	Gln	Gln	Ile		
			115				120						125				
Ile	Val	Ser	Leu	Thr	Glu	Ser	Cys	Arg	Ser	Cys	Ser	His	Ala	Leu	Ala		
			130				135					140					
Ala	His	Val	Ser	His	Leu	Glu	Asn	Val	Ser	Glu	Glu	Glu	Met	Asn	Arg		
145						150				155					160		
Leu	Leu	Gly	Ile	Val	Leu	Asp	Val	Glu	Tyr	Leu	Phe	Thr	Cys	Val	His		
				165						170					175		
Lys	Glu	Glu	Asp	Ala	Asp	Thr	Lys	Gln	Val	Tyr	Phe	Tyr	Leu	Phe	Lys		
			180					185					190				
Leu	Leu	Arg	Lys	Ser	Ile	Leu	Gln	Arg	Gly	Lys	Pro	Val	Val	Glu	Gly		
			195				200					205					
Ser	Leu	Glu	Lys	Lys	Pro	Pro	Phe	Glu	Lys	Pro	Ser	Ile	Glu	Gln	Gly		
			210				215					220					
Val	Asn	Asn	Phe	Val	Gln	Tyr	Lys	Phe	Ser	His	Leu	Pro	Ala	Lys	Glu		
225					230					235					240		
Arg	Gln	Thr	Ile	Val	Glu	Leu	Ala	Lys	Met	Phe	Leu	Asn	Arg	Ile	Asn		
				245					250					255			
Tyr	Trp	His	Leu	Glu	Ala	Pro	Ser	Gln	Arg	Arg	Leu	Arg	Ser	Pro	Asn		
			260				265						270				
Asp	Asp	Ile	Ser	Gly	Tyr	Lys	Glu	Asn	Tyr	Thr	Arg	Trp	Leu	Cys	Tyr		
			275				280					285					
Cys	Asn	Val	Pro	Gln	Phe	Cys	Asp	Ser	Leu	Pro	Arg	Tyr	Glu	Thr	Thr		
			290				295				300						
Gln	Val	Phe	Gly	Arg	Thr	Leu	Leu	Arg	Ser	Val	Phe	Thr	Val	Met	Arg		
305					310					315					320		
Arg	Gln	Leu	Leu	Glu	Gln	Ala	Arg	Gln	Glu	Lys	Asp	Lys	Leu	Pro	Leu		
				325					330					335			
Glu	Lys	Arg	Thr	Leu	Ile	Leu	Thr	His	Phe	Pro	Lys	Phe	Leu	Ser	Met		
			340					345					350				
Leu	Glu	Glu	Glu	Val	Tyr	Ser	Gln	Asn	Ser	Pro	Ile	Trp	Asp	Gln	Asp		
			355				360					365					
Phe	Leu	Ser	Ala	Ser	Ser	Arg	Thr	Ser	Gln	Leu	Gly	Ile	Gln	Thr	Val		
			370				375					380					
Ile	Asn	Pro	Pro	Pro	Val	Ala	Gly	Thr	Ile	Ser	Tyr	Asn	Ser	Thr	Ser		
385					390					395					400		
Ser	Ser	Leu	Glu	Gln	Pro	Asn	Ala	Gly	Ser	Ser	Ser	Pro	Ala	Cys	Lys		
				405					410					415			
Ala	Ser	Ser	Gly	Leu	Glu	Ala	Asn	Pro	Gly	Glu	Lys	Arg	Lys	Met	Thr		
			420					425					430				
Asp	Ser	His	Val	Leu	Glu	Glu	Ala	Lys	Lys	Pro	Arg	Val	Met	Gly	Asp		
			435				440					445					
Ile	Pro	Met	Glu	Leu	Ile	Asn	Glu	Val	Met	Ser	Thr	Ile	Thr	Asp	Pro		
			450			455					460						
Ala	Ala	Met	Leu	Gly	Pro	Glu	Thr	Asn	Phe	Leu	Ser	Ala	His	Ser	Ala		
465					470					475					480		

Arg Asp Glu Ala Ala Arg Leu Glu Glu Arg Arg Gly Val Ile Glu Phe  
 485 490 495  
 His Val Val Gly Asn Ser Leu Asn Gln Lys Pro Asn Lys Lys Ile Leu  
 500 505 510  
 Met Trp Leu Val Gly Leu Gln Asn Val Phe Ser His Gln Leu Pro Arg  
 515 520 525  
 Met Pro Lys Glu Tyr Ile Thr Arg Leu Val Phe Asp Pro Lys His Lys  
 530 535 540  
 Thr Leu Ala Leu Ile Lys Asp Gly Arg Val Ile Gly Gly Ile Cys Phe  
 545 550 555 560  
 Arg Met Phe Pro Ser Gln Gly Phe Thr Glu Ile Val Phe Cys Ala Val  
 565 570 575  
 Thr Ser Asn Glu Gln Val Lys Gly Tyr Gly Thr His Leu Met Asn His  
 580 585 590  
 Leu Lys Glu Tyr His Ile Lys His Asp Ile Leu Asn Phe Leu Thr Tyr  
 595 600 605  
 Ala Asp Glu Tyr Ala Ile Gly Tyr Phe Lys Lys Gln Gly Phe Ser Lys  
 610 615 620  
 Glu Ile Lys Ile Pro Lys Thr Lys Tyr Val Gly Tyr Ile Lys Asp Tyr  
 625 630 635 640  
 Glu Gly Ala Thr Leu Met Gly Cys Glu Leu Asn Pro Arg Ile Pro Tyr  
 645 650 655  
 Thr Glu Phe Ser Val Ile Ile Lys Lys Gln Lys Glu Ile Ile Lys Lys  
 660 665 670  
 Leu Ile Glu Arg Lys Gln Ala Gln Ile Arg Lys Val Tyr Pro Gly Leu  
 675 680 685  
 Ser Cys Phe Lys Asp Gly Val Arg Gln Ile Pro Ile Glu Ser Ile Pro  
 690 695 700  
 Gly Ile Arg Glu Thr Gly Trp Lys Pro Ser Gly Lys Glu Lys Ser Lys  
 705 710 715 720  
 Glu Pro Arg Asp Pro Asp Gln Leu Tyr Ser Thr Leu Lys Ser Ile Leu  
 725 730 735  
 Gln Gln Val Lys Ser His Gln Ser Ala Trp Pro Phe Met Glu Pro Val  
 740 745 750  
 Lys Arg Thr Glu Ala Pro Gly Tyr Tyr Glu Val Ile Arg Ser Pro Met  
 755 760 765  
 Asp Leu Lys Thr Met Ser Glu Arg Leu Lys Asn Arg Tyr Tyr Val Ser  
 770 775 780  
 Lys Lys Leu Phe Met Ala Asp Leu Gln Arg Val Phe Thr Asn Cys Lys  
 785 790 795 800  
 Glu Tyr Asn Ala Pro Glu Ser Glu Tyr Tyr Lys Cys Ala Asn Ile Leu  
 805 810 815  
 Glu Lys Phe Phe Phe Ser Lys Ile Lys Glu Ala Gly Leu Ile Asp Lys  
 820 825 830

## (2) INFORMATION FOR SEQ ID NO:2:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 481 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: None

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Leu Glu Glu Glu Val Tyr Ser Gln Asn Ser Pro Ile Trp Asp Gln  
 1 5 10 15  
 Asp Phe Leu Ser Ala Ser Ser Arg Thr Ser Gln Leu Gly Ile Gln Thr  
 20 25 30  
 Val Ile Asn Pro Pro Pro Val Ala Gly Thr Ile Ser Tyr Asn Ser Thr  
 35 40 45

Ser	Ser	Ser	Leu	Glu	Gln	Pro	Asn	Ala	Gly	Ser	Ser	Ser	Pro	Ala	Cys
50						55				60					
Lys	Ala	Ser	Ser	Gly	Leu	Glu	Ala	Asn	Pro	Gly	Glu	Lys	Arg	Lys	Met
65				70						75					80
Thr	Asp	Ser	His	Val	Leu	Glu	Glu	Ala	Lys	Lys	Pro	Arg	Val	Met	Gly
			85						90					95	
Asp	Ile	Pro	Met	Glu	Leu	Ile	Asn	Glu	Val	Met	Ser	Thr	Ile	Thr	Asp
			100					105					110		
Pro	Ala	Ala	Met	Leu	Gly	Pro	Glu	Thr	Asn	Phe	Leu	Ser	Ala	His	Ser
			115					120					125		
Ala	Arg	Asp	Glu	Ala	Ala	Arg	Leu	Glu	Glu	Arg	Arg	Gly	Val	Ile	Glu
			130				135				140				
Phe	His	Val	Val	Gly	Asn	Ser	Leu	Asn	Gln	Lys	Pro	Asn	Lys	Lys	Ile
145				150						155					160
Leu	Met	Trp	Leu	Val	Gly	Leu	Gln	Asn	Val	Phe	Ser	His	Gln	Leu	Pro
			165						170					175	
Arg	Met	Pro	Lys	Glu	Tyr	Ile	Thr	Arg	Leu	Val	Phe	Asp	Pro	Lys	His
			180					185					190		
Lys	Thr	Leu	Ala	Leu	Ile	Lys	Asp	Gly	Arg	Val	Ile	Gly	Gly	Ile	Cys
			195				200					205			
Phe	Arg	Met	Phe	Pro	Ser	Gln	Gly	Phe	Thr	Glu	Ile	Val	Phe	Cys	Ala
			210			215					220				
Val	Thr	Ser	Asn	Glu	Gln	Val	Lys	Gly	Tyr	Gly	Thr	His	Leu	Met	Asn
225				230						235					240
His	Leu	Lys	Glu	Tyr	His	Ile	Lys	His	Asp	Ile	Leu	Asn	Phe	Leu	Thr
			245						250					255	
Tyr	Ala	Asp	Glu	Tyr	Ala	Ile	Gly	Tyr	Phe	Lys	Lys	Gln	Gly	Phe	Ser
			260					265					270		
Lys	Glu	Ile	Lys	Ile	Pro	Lys	Thr	Lys	Tyr	Val	Gly	Tyr	Ile	Lys	Asp
			275				280					285			
Tyr	Glu	Gly	Ala	Thr	Leu	Met	Gly	Cys	Glu	Leu	Asn	Pro	Arg	Ile	Pro
			290			295					300				
Tyr	Thr	Glu	Phe	Ser	Val	Ile	Ile	Lys	Lys	Gln	Lys	Glu	Ile	Ile	Lys
305				310						315					320
Lys	Leu	Ile	Glu	Arg	Lys	Gln	Ala	Gln	Ile	Arg	Lys	Val	Tyr	Pro	Gly
			325						330					335	
Leu	Ser	Cys	Phe	Lys	Asp	Gly	Val	Arg	Gln	Ile	Pro	Ile	Glu	Ser	Ile
			340					345					350		
Pro	Gly	Ile	Arg	Glu	Thr	Gly	Trp	Lys	Pro	Ser	Gly	Lys	Glu	Lys	Ser
			355				360					365			
Lys	Glu	Pro	Arg	Asp	Pro	Asp	Gln	Leu	Tyr	Ser	Thr	Leu	Lys	Ser	Ile
			370			375					380				
Leu	Gln	Gln	Val	Lys	Ser	His	Gln	Ser	Ala	Trp	Pro	Phe	Met	Glu	Pro
385				390						395					400
Val	Lys	Arg	Thr	Glu	Ala	Pro	Gly	Tyr	Tyr	Glu	Val	Ile	Arg	Ser	Pro
			405						410					415	
Met	Asp	Leu	Lys	Thr	Met	Ser	Glu	Arg	Leu	Lys	Asn	Arg	Tyr	Tyr	Val
			420					425					430		
Ser	Lys	Lys	Leu	Phe	Met	Ala	Asp	Leu	Gln	Arg	Val	Phe	Thr	Asn	Cys
			435				440					445			
Lys	Glu	Tyr	Asn	Ala	Pro	Glu	Ser	Glu	Tyr	Tyr	Lys	Cys	Ala	Asn	Ile
			450			455					460				
Leu	Glu	Lys	Phe	Phe	Phe	Ser	Lys	Ile	Lys	Glu	Ala	Gly	Leu	Ile	Asp
465				470						475					480
Lys															

## (2) INFORMATION FOR SEQ ID NO:3:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 203 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```

Arg Val Val Gln His Thr Lys Gly Cys Lys Arg Lys Thr Asn Gly Gly
 1          5          10          15
Cys Pro Ile Cys Lys Gln Leu Ile Ala Leu Cys Cys Tyr His Ala Lys
 20          25          30
His Cys Gln Glu Asn Lys Cys Pro Val Pro Phe Cys Leu Asn Ile Lys
 35          40          45
Gln Lys Leu Arg Gln Gln Gln Leu Gln His Arg Leu Gln Gln Ala Gln
 50          55          60
Met Leu Arg Arg Arg Met Ala Ser Met Arg Thr Gly Val Val Gly Gln
 65          70          75          80
Gln Gln Gly Leu Pro Ser Pro Thr Pro Ala Thr Pro Thr Thr Pro Thr
 85          90          95
Gly Gln Gln Pro Thr Thr Pro Gln Thr Pro Gln Pro Thr Ser Gln Pro
100          105          110
Gln Pro Thr Pro Pro Asn Ser Met Pro Pro Tyr Leu Pro Arg Thr Gln
115          120          125
Ala Ala Gly Pro Val Ser Gln Gly Lys Ala Ala Gly Gln Val Thr Pro
130          135          140
Pro Thr Pro Pro Gln Thr Ala Gln Pro Pro Leu Pro Gly Pro Pro Pro
145          150          155          160
Thr Ala Val Glu Met Ala Met Gln Ile Gln Arg Ala Ala Glu Thr Gln
165          170          175          180
Arg Gln Met Ala His Val Gln Ile Phe Gln Arg Pro Ile Gln His Gln
185          190
Met Pro Pro Met Thr Pro Met Ala Pro Met Gly
195          200

```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 351 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```

Met Ser Glu Ala Gly Gly Ala Gly Pro Gly Gly Cys Gly Ala Gly Ala
 1          5          10          15
Gly Ala Gly Ala Gly Pro Gly Ala Leu Pro Pro Gln Pro Ala Ala Leu
 20          25          30
Pro Pro Ala Pro Pro Gln Gly Ser Pro Cys Ala Ala Ala Gly Gly
 35          40          45
Ser Gly Ala Cys Gly Pro Ala Thr Ala Val Ala Ala Ala Gly Thr Ala
 50          55          60
Glu Gly Pro Gly Gly Gly Gly Ser Ala Arg Ile Ala Val Lys Lys Ala
 65          70          75          80
Gln Leu Arg Ser Ala Pro Arg Ala Lys Lys Leu Glu Lys Leu Gly Val
 85          90          95
Tyr Ser Ala Cys Lys Ala Glu Glu Ser Cys Lys Cys Asn Gly Trp Lys
100          105          110
Asn Pro Asn Pro Ser Pro Thr Pro Pro Arg Ala Asp Leu Gln Gln Ile
115          120          125
Ile Val Ser Leu Thr Glu Ser Cys Arg Ser Cys Ser His Ala Leu Ala
130          135          140

```

Ala	His	Val	Ser	His	Leu	Glu	Asn	Val	Ser	Glu	Glu	Glu	Met	Asn	Arg
145					150					155				160	
Leu	Leu	Gly	Ile	Val	Leu	Asp	Val	Glu	Tyr	Leu	Phe	Thr	Cys	Val	His
			165						170					175	
Lys	Glu	Glu	Asp	Ala	Asp	Thr	Lys	Gln	Val	Tyr	Phe	Tyr	Leu	Phe	Lys
			180					185					190		
Leu	Leu	Arg	Lys	Ser	Ile	Leu	Gln	Arg	Gly	Lys	Pro	Val	Val	Glu	Gly
		195					200					205			
Ser	Leu	Glu	Lys	Lys	Pro	Pro	Phe	Glu	Lys	Pro	Ser	Ile	Glu	Gln	Gly
	210					215					220				
Val	Asn	Asn	Phe	Val	Gln	Tyr	Lys	Phe	Ser	His	Leu	Pro	Ala	Lys	Glu
225					230					235					240
Arg	Gln	Thr	Ile	Val	Glu	Leu	Ala	Lys	Met	Phe	Leu	Asn	Arg	Ile	Asn
			245						250					255	
Tyr	Trp	His	Leu	Glu	Ala	Pro	Ser	Gln	Arg	Arg	Leu	Arg	Ser	Pro	Asn
			260					265					270		
Asp	Asp	Ile	Ser	Gly	Tyr	Lys	Glu	Asn	Tyr	Thr	Arg	Trp	Leu	Cys	Tyr
	275						280					285			
Cys	Asn	Val	Pro	Gln	Phe	Cys	Asp	Ser	Leu	Pro	Arg	Tyr	Glu	Thr	Thr
	290					295					300				
Gln	Val	Phe	Gly	Arg	Thr	Leu	Leu	Arg	Ser	Val	Phe	Thr	Val	Met	Arg
305					310					315					320
Arg	Gln	Leu	Leu	Glu	Gln	Ala	Arg	Gln	Glu	Lys	Asp	Lys	Leu	Pro	Leu
			325						330					335	
Glu	Lys	Arg	Thr	Leu	Ile	Leu	Thr	His	Phe	Pro	Lys	Phe	Leu	Ser	
			340					345					350		

## (2) INFORMATION FOR SEQ ID NO:5:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 476 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: None

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Met	Leu	Glu	Glu	Glu	Ile	Tyr	Gly	Ala	Asn	Ser	Pro	Ile	Trp	Glu	Ser
1				5					10					15	
Gly	Phe	Thr	Met	Pro	Pro	Ser	Glu	Gly	Thr	Gln	Leu	Val	Pro	Arg	Pro
			20					25					30		
Ala	Ser	Val	Ser	Ala	Ala	Val	Val	Pro	Ser	Thr	Pro	Ile	Phe	Ser	Pro
		35					40					45			
Ser	Met	Gly	Gly	Gly	Ser	Asn	Ser	Ser	Leu	Ser	Leu	Asp	Ser	Ala	Gly
	50					55					60				
Ala	Glu	Pro	Met	Pro	Gly	Glu	Lys	Arg	Thr	Leu	Pro	Glu	Asn	Leu	Thr
65					70					75				80	
Leu	Glu	Asp	Ala	Lys	Arg	Leu	Arg	Val	Met	Gly	Asp	Ile	Pro	Met	Glu
			85						90				95		
Leu	Val	Asn	Glu	Val	Met	Leu	Thr	Ile	Thr	Asp	Pro	Ala	Ala	Met	Leu
			100						105				110		
Gly	Pro	Glu	Thr	Ser	Leu	Leu	Ser	Ala	Asn	Ala	Ala	Arg	Asp	Glu	Thr
		115					120					125			
Ala	Arg	Leu	Glu	Glu	Arg	Arg	Gly	Ile	Ile	Glu	Phe	His	Val	Ile	Gly
	130					135				140					
Asn	Ser	Leu	Thr	Pro	Lys	Ala	Asn	Arg	Arg	Val	Leu	Leu	Trp	Leu	Val
145					150					155					160
Gly	Leu	Gln	Asn	Val	Phe	Ser	His	Gln	Leu	Pro	Arg	Met	Pro	Lys	Glu
			165						170					175	
Tyr	Ile	Ala	Arg	Leu	Val	Phe	Asp	Pro	Lys	His	Lys	Thr	Leu	Ala	Leu
			180					185					190		

Ile Lys Asp Gly Arg Val Ile Gly Gly Ile Cys Phe Arg Met Phe Pro  
 195 200 205  
 Thr Gln Gly Phe Thr Glu Ile Val Phe Cys Ala Val Thr Ser Asn Glu  
 210 215 220  
 Gln Val Lys Gly Tyr Gly Thr His Leu Met Asn His Leu Lys Glu Tyr  
 225 230 235 240  
 His Ile Lys His Asn Ile Leu Tyr Phe Leu Thr Tyr Ala Asp Glu Tyr  
 245 250 255  
 Ala Ile Gly Tyr Phe Lys Lys Gln Gly Phe Ser Lys Asp Ile Lys Val  
 260 265 270  
 Pro Lys Ser Arg Tyr Leu Gly Tyr Ile Lys Asp Tyr Glu Gly Ala Thr  
 275 280 285  
 Leu Met Glu Cys Glu Leu Asn Pro Arg Ile Pro Tyr Thr Glu Leu Ser  
 290 295 300  
 His Ile Ile Lys Lys Gln Lys Glu Ile Ile Lys Lys Leu Ile Glu Arg  
 305 310 315 320  
 Lys Gln Ala Gln Ile Arg Lys Val Tyr Pro Gly Leu Ser Cys Phe Lys  
 325 330 335  
 Glu Gly Val Arg Gln Ile Pro Val Glu Ser Val Pro Gly Ile Arg Glu  
 340 345 350  
 Thr Gly Trp Lys Pro Leu Gly Lys Glu Lys Gly Lys Glu Leu Lys Asp  
 355 360 365  
 Pro Asp Gln Leu Tyr Thr Thr Leu Lys Asn Leu Leu Ala Gln Ile Lys  
 370 375 380  
 Ser His Pro Ser Ala Trp Pro Phe Met Glu Pro Val Lys Lys Ser Glu  
 385 390 395 400  
 Ala Pro Asp Tyr Tyr Glu Val Ile Arg Phe Pro Ile Asp Leu Lys Thr  
 405 410 415  
 Met Thr Glu Arg Leu Arg Ser Arg Tyr Tyr Val Thr Arg Lys Leu Phe  
 420 425 430  
 Val Ala Asp Leu Gln Arg Val Ile Ala Asn Cys Arg Glu Tyr Asn Pro  
 435 440 445  
 Pro Asp Ser Glu Tyr Cys Arg Cys Ala Ser Ala Leu Glu Lys Phe Phe  
 450 455 460  
 Tyr Phe Lys Leu Lys Glu Gly Gly Leu Ile Asp Lys  
 465 470 475

## (2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2414 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Met Ala Glu Asn Val Val Glu Pro Gly Pro Pro Ser Ala Lys Arg Pro  
 1 5 10 15  
 Lys Leu Ser Ser Pro Ala Leu Ser Ala Ser Ala Ser Asp Gly Thr Asp  
 20 25 30  
 Phe Gly Ser Leu Phe Asp Leu Glu His Asp Leu Pro Asp Glu Leu Ile  
 35 40 45  
 Asn Ser Thr Glu Leu Gly Leu Thr Asn Gly Gly Asp Ile Asn Gln Leu  
 50 55 60  
 Gln Thr Ser Leu Gly Met Val Gln Asp Ala Ala Ser Lys His Lys Gln  
 65 70 75 80  
 Leu Ser Glu Leu Leu Arg Ser Gly Ser Ser Pro Asn Leu Asn Met Gly  
 85 90 95  
 Val Gly Gly Pro Gly Gln Val Met Ala Ser Gln Ala Gln Gln Ser Ser  
 100 105 110

Pro	Gly	Leu	Gly	Leu	Ile	Asn	Ser	Met	Val	Lys	Ser	Pro	Met	Thr	Gln
		115					120					125			
Ala	Gly	Leu	Thr	Ser	Pro	Asn	Met	Gly	Met	Gly	Thr	Ser	Gly	Pro	Asn
	130					135					140				
Gln	Gly	Pro	Thr	Gln	Ser	Thr	Gly	Met	Met	Asn	Ser	Pro	Val	Asn	Gln
145					150					155					160
Pro	Ala	Met	Gly	Met	Asn	Thr	Gly	Thr	Asn	Ala	Gly	Met	Asn	Pro	Gly
			165						170					175	
Met	Leu	Ala	Ala	Gly	Asn	Gly	Gln	Gly	Ile	Met	Pro	Asn	Gln	Val	Met
		180						185					190		
Asn	Gly	Ser	Ile	Gly	Ala	Gly	Arg	Gly	Arg	Gln	Asp	Met	Gln	Tyr	Pro
	195						200					205			
Asn	Pro	Gly	Met	Gly	Ser	Ala	Gly	Asn	Leu	Leu	Thr	Glu	Pro	Leu	Gln
	210					215					220				
Gln	Gly	Ser	Pro	Gln	Met	Gly	Gly	Gln	Thr	Gly	Leu	Arg	Gly	Pro	Gln
225					230					235					240
Pro	Leu	Lys	Met	Gly	Met	Met	Asn	Asn	Pro	Asn	Pro	Tyr	Gly	Ser	Pro
			245						250					255	
Tyr	Thr	Gln	Asn	Pro	Gly	Gln	Gln	Ile	Gly	Ala	Ser	Gly	Leu	Gly	Leu
		260						265					270		
Gln	Ile	Gln	Thr	Lys	Thr	Val	Leu	Ser	Asn	Asn	Leu	Ser	Pro	Phe	Ala
	275						280					285			
Met	Asp	Lys	Lys	Ala	Val	Pro	Gly	Gly	Gly	Met	Pro	Asn	Met	Gly	Gln
	290					295					300				
Gln	Pro	Ala	Pro	Gln	Val	Gln	Gln	Pro	Gly	Leu	Val	Thr	Pro	Val	Ala
305					310					315					320
Gln	Gly	Met	Gly	Ser	Gly	Ala	His	Thr	Ala	Asp	Pro	Glu	Lys	Arg	Lys
			325						330					335	
Leu	Ile	Gln	Gln	Gln	Leu	Val	Leu	Leu	His	Ala	His	Lys	Cys	Gln	
	340						345					350			
Arg	Arg	Glu	Gln	Ala	Asn	Gly	Glu	Val	Arg	Gln	Cys	Asn	Leu	Pro	His
	355						360					365			
Cys	Arg	Thr	Met	Lys	Asn	Val	Leu	Asn	His	Met	Thr	His	Cys	Gln	Ser
	370				375						380				
Gly	Lys	Ser	Cys	Gln	Val	Ala	His	Cys	Ala	Ser	Ser	Arg	Gln	Ile	Ile
385					390					395					400
Ser	His	Trp	Lys	Asn	Cys	Thr	Arg	His	Asp	Cys	Pro	Val	Cys	Leu	Pro
			405						410					415	
Leu	Lys	Asn	Ala	Gly	Asp	Lys	Arg	Asn	Gln	Gln	Pro	Ile	Leu	Thr	Gly
		420						425					430		
Ala	Pro	Val	Gly	Leu	Gly	Asn	Pro	Ser	Ser	Leu	Gly	Val	Gly	Gln	Gln
	435						440					445			
Ser	Ala	Pro	Asn	Leu	Ser	Thr	Val	Ser	Gln	Ile	Asp	Pro	Ser	Ser	Ile
	450					455					460				
Glu	Arg	Ala	Tyr	Ala	Ala	Leu	Gly	Leu	Pro	Tyr	Gln	Val	Asn	Gln	Met
465					470					475					480
Pro	Thr	Gln	Pro	Gln	Val	Gln	Ala	Lys	Asn	Gln	Gln	Asn	Gln	Gln	Pro
				485					490					495	
Gly	Gln	Ser	Pro	Gln	Gly	Met	Arg	Pro	Met	Ser	Asn	Met	Ser	Ala	Ser
		500						505					510		
Pro	Met	Gly	Val	Asn	Gly	Gly	Val	Gly	Val	Gln	Thr	Pro	Ser	Leu	Leu
	515						520					525			
Ser	Asp	Ser	Met	Leu	His	Ser	Ala	Ile	Asn	Ser	Gln	Asn	Pro	Met	Met
	530					535					540				
Ser	Glu	Asn	Ala	Ser	Val	Pro	Ser	Leu	Gly	Pro	Met	Pro	Thr	Ala	Ala
545					550					555					560
Gln	Pro	Ser	Thr	Thr	Gly	Ile	Arg	Lys	Gln	Trp	His	Glu	Asp	Ile	Thr
			565						570					575	
Gln	Asp	Leu	Arg	Asn	His	Leu	Val	His	Lys	Leu	Val	Gln	Ala	Ile	Phe
		580						585					590		
Pro	Thr	Pro	Asp	Pro	Ala	Ala	Leu	Lys	Asp	Arg	Arg	Met	Glu	Asn	Leu
		595					600					605			

Val Ala Tyr Ala Arg Lys Val Glu Gly Asp Met Tyr Glu Ser Ala Asn  
 610 615 620  
 Asn Arg Ala Glu Tyr Tyr His Leu Leu Ala Glu Lys Ile Tyr Lys Ile  
 625 630 635 640  
 Gln Lys Glu Leu Glu Glu Lys Arg Arg Thr Arg Leu Gln Lys Gln Asn  
 645 650 655  
 Met Leu Pro Asn Ala Ala Gly Met Val Pro Val Ser Met Asn Pro Gly  
 660 665 670  
 Pro Asn Met Gly Gln Pro Gln Pro Gly Met Thr Ser Asn Gly Pro Leu  
 675 680 685  
 Pro Asp Pro Ser Met Ile Arg Gly Ser Val Pro Asn Gln Met Met Pro  
 690 695 700  
 Arg Ile Thr Pro Gln Ser Gly Leu Asn Gln Phe Gly Gln Met Ser Met  
 705 710 715 720  
 Ala Gln Pro Pro Ile Val Pro Arg Gln Thr Pro Pro Leu Gln His His  
 725 730 735  
 Gly Gln Leu Ala Gln Pro Gly Ala Leu Asn Pro Pro Met Gly Tyr Gly  
 740 745 750  
 Pro Arg Met Gln Gln Pro Ser Asn Gln Gly Gln Phe Leu Pro Gln Thr  
 755 760 765  
 Gln Phe Pro Ser Gln Gly Met Asn Val Thr Asn Ile Pro Leu Ala Pro  
 770 775 780  
 Ser Ser Gly Gln Ala Pro Val Ser Gln Ala Gln Met Ser Ser Ser Ser  
 785 790 795 800  
 Cys Pro Val Asn Ser Pro Ile Met Pro Pro Gly Ser Gln Gly Ser His  
 805 810 815  
 Ile His Cys Pro Gln Leu Pro Gln Pro Ala Leu His Gln Asn Ser Pro  
 820 825 830  
 Ser Pro Val Pro Ser Arg Thr Pro Thr Pro His His Thr Pro Pro Ser  
 835 840 845  
 Ile Gly Ala Gln Gln Pro Pro Ala Thr Thr Ile Pro Ala Pro Val Pro  
 850 855 860  
 Thr Pro Pro Ala Met Pro Pro Gly Pro Gln Ser Gln Ala Leu His Pro  
 865 870 875 880  
 Pro Pro Arg Gln Thr Pro Thr Pro Pro Thr Thr Gln Leu Pro Gln Gln  
 885 890 895  
 Val Gln Pro Ser Leu Pro Ala Ala Pro Ser Ala Asp Gln Pro Gln Gln  
 900 905 910  
 Gln Pro Arg Ser Gln Gln Ser Thr Ala Ala Ser Val Pro Thr Pro Asn  
 915 920 925  
 Ala Pro Leu Leu Pro Pro Gln Pro Ala Thr Pro Leu Ser Gln Pro Ala  
 930 935 940  
 Val Ser Ile Glu Gly Gln Val Ser Asn Pro Pro Ser Thr Ser Ser Thr  
 945 950 955 960  
 Glu Val Asn Ser Gln Ala Ile Ala Glu Lys Gln Pro Ser Gln Glu Val  
 965 970 975  
 Lys Met Glu Ala Lys Met Glu Val Asp Gln Pro Glu Pro Ala Asp Thr  
 980 985 990  
 Gln Pro Glu Asp Ile Ser Glu Ser Lys Val Glu Asp Cys Lys Met Glu  
 995 1000 1005  
 Ser Thr Glu Thr Glu Glu Arg Ser Thr Glu Leu Lys Thr Glu Ile Lys  
 1010 1015 1020  
 Glu Glu Glu Asp Gln Pro Ser Thr Ser Ala Thr Gln Ser Ser Pro Ala  
 1025 1030 1035 1040  
 Pro Gly Gln Ser Lys Lys Lys Ile Phe Lys Pro Glu Glu Leu Arg Gln  
 1045 1050 1055  
 Ala Leu Met Pro Thr Leu Glu Ala Leu Tyr Arg Gln Asp Pro Glu Ser  
 1060 1065 1070  
 Leu Pro Phe Arg Gln Pro Val Asp Pro Gln Leu Leu Gly Ile Pro Asp  
 1075 1080 1085  
 Tyr Phe Asp Ile Val Lys Ser Pro Met Asp Leu Ser Thr Ile Lys Arg  
 1090 1095 1100

Lys Leu Asp Thr Gly Gln Tyr Gln Glu Pro Trp Gln Tyr Val Asp Asp  
 105 1110 1115 1120  
 Ile Trp Leu Met Phe Asn Asn Ala Trp Leu Tyr Asn Arg Lys Thr Ser  
 1125 1130 1135  
 Arg Val Tyr Lys Tyr Cys Ser Lys Leu Ser Glu Val Phe Glu Gln Glu  
 1140 1145 1150  
 Ile Asp Pro Val Met Gln Ser Leu Gly Tyr Cys Cys Gly Arg Lys Leu  
 1155 1160 1165  
 Glu Phe Ser Pro Gln Thr Leu Cys Cys Tyr Gly Lys Gln Leu Cys Thr  
 1170 1175 1180  
 Ile Pro Arg Asp Ala Thr Tyr Tyr Ser Tyr Gln Asn Arg Tyr His Phe  
 185 1190 1195 1200  
 Cys Glu Lys Cys Phe Asn Glu Ile Gln Gly Glu Ser Val Ser Leu Gly  
 1205 1210 1215  
 Asp Asp Pro Ser Gln Pro Gln Thr Thr Ile Asn Lys Glu Gln Phe Ser  
 1220 1225 1230  
 Lys Arg Lys Asn Asp Thr Leu Asp Pro Glu Leu Phe Val Glu Cys Thr  
 1235 1240 1245  
 Glu Cys Gly Arg Lys Met His Gln Ile Cys Val Leu His His Glu Ile  
 1250 1255 1260  
 Ile Trp Pro Ala Gly Phe Val Cys Asp Gly Cys Leu Lys Lys Ser Ala  
 265 1270 1275 1280  
 Arg Thr Arg Lys Glu Asn Lys Phe Ser Ala Lys Arg Leu Pro Ser Thr  
 1285 1290 1295  
 Arg Leu Gly Thr Phe Leu Glu Asn Arg Val Asn Asp Phe Leu Arg Arg  
 1300 1305 1310  
 Gln Asn His Pro Glu Ser Gly Glu Val Thr Val Arg Val Val His Ala  
 1315 1320 1325  
 Ser Asp Lys Thr Val Glu Val Lys Pro Gly Met Lys Ala Arg Phe Val  
 1330 1335 1340  
 Asp Ser Gly Glu Met Ala Glu Ser Phe Pro Tyr Arg Thr Lys Ala Leu  
 345 1350 1355 1360  
 Phe Ala Phe Glu Glu Ile Asp Gly Val Asp Leu Cys Phe Phe Gly Met  
 1365 1370 1375  
 His Val Gln Glu Tyr Gly Ser Asp Cys Pro Pro Pro Asn Gln Arg Arg  
 1380 1385 1390  
 Val Tyr Ile Ser Tyr Leu Asp Ser Val His Phe Phe Arg Pro Lys Cys  
 1395 1400 1405  
 Leu Arg Thr Ala Val Tyr His Glu Ile Leu Ile Gly Tyr Leu Glu Tyr  
 1410 1415 1420  
 Val Lys Lys Leu Gly Tyr Thr Thr Gly His Ile Trp Ala Cys Pro Pro  
 425 1430 1435 1440  
 Ser Glu Gly Asp Asp Tyr Ile Phe His Cys His Pro Pro Asp Gln Lys  
 1445 1450 1455  
 Ile Pro Lys Pro Lys Arg Leu Gln Glu Trp Tyr Lys Lys Met Leu Asp  
 1460 1465 1470  
 Lys Ala Val Ser Glu Arg Ile Val His Asp Tyr Lys Asp Ile Phe Lys  
 1475 1480 1485  
 Gln Ala Thr Glu Asp Arg Leu Thr Ser Ala Lys Glu Leu Pro Tyr Phe  
 1490 1495 1500  
 Glu Gly Asp Phe Trp Pro Asn Val Leu Glu Glu Ser Ile Lys Glu Leu  
 505 1510 1515 1520  
 Glu Gln Glu Glu Glu Arg Lys Arg Glu Glu Asn Thr Ser Asn Glu  
 1525 1530 1535  
 Ser Thr Asp Val Thr Lys Gly Asp Ser Lys Asn Ala Lys Lys Lys Asn  
 1540 1545 1550  
 Asn Lys Lys Thr Ser Lys Asn Lys Ser Ser Leu Ser Arg Gly Asn Lys  
 1555 1560 1565  
 Lys Lys Pro Gly Met Pro Asn Val Ser Asn Asp Leu Ser Gln Lys Leu  
 1570 1575 1580  
 Tyr Ala Thr Met Glu Lys His Lys Glu Val Phe Phe Val Ile Arg Leu  
 585 1590 1595 1600

Ile Ala Gly Pro Ala Ala Asn Ser Leu Pro Pro Ile Val Asp Pro Asp  
 1605 1610 1615  
 Pro Leu Ile Pro Cys Asp Leu Met Asp Gly Arg Asp Ala Phe Leu Thr  
 1620 1625 1630  
 Leu Ala Arg Asp Lys His Leu Glu Phe Ser Ser Leu Arg Arg Ala Gln  
 1635 1640 1645  
 Trp Ser Thr Met Cys Met Leu Val Glu Leu His Thr Gln Ser Gln Asp  
 1650 1655 1660  
 Arg Phe Val Tyr Thr Cys Asn Glu Cys Lys His His Val Glu Thr Arg  
 665 1670 1675 1680  
 Trp His Cys Thr Val Cys Glu Asp Tyr Asp Leu Cys Ile Thr Cys Tyr  
 1685 1690 1695  
 Asn Thr Lys Asn His Asp His Lys Met Glu Lys Leu Gly Leu Gly Leu  
 1700 1705 1710  
 Asp Asp Glu Ser Asn Asn Gln Gln Ala Ala Thr Gln Ser Pro Gly  
 1715 1720 1725  
 Asp Ser Arg Arg Leu Ser Ile Gln Arg Cys Ile Gln Ser Leu Val His  
 1730 1735 1740  
 Ala Cys Gln Cys Arg Asn Ala Asn Cys Ser Leu Pro Ser Cys Gln Lys  
 745 1750 1755 1760  
 Met Lys Arg Val Val Gln His Thr Lys Gly Cys Lys Arg Lys Thr Asn  
 1765 1770 1775  
 Gly Gly Cys Pro Ile Cys Lys Gln Leu Ile Ala Leu Cys Cys Tyr His  
 1780 1785 1790  
 Ala Lys His Cys Gln Glu Asn Lys Cys Pro Val Pro Phe Cys Leu Asn  
 1795 1800 1805  
 Ile Lys Gln Lys Leu Arg Gln Gln Gln Leu Gln His Arg Leu Gln Gln  
 1810 1815 1820  
 Ala Gln Met Leu Arg Arg Arg Met Ala Ser Met Gln Arg Thr Gly Val  
 825 1830 1835 1840  
 Val Gly Gln Gln Gln Gly Leu Pro Ser Pro Thr Pro Ala Thr Pro Thr  
 1845 1850 1855  
 Thr Pro Thr Gly Gln Gln Pro Thr Thr Pro Gln Thr Pro Gln Pro Thr  
 1860 1865 1870  
 Ser Gln Pro Gln Pro Thr Pro Pro Asn Ser Met Pro Pro Tyr Leu Pro  
 1875 1880 1885  
 Arg Thr Gln Ala Ala Gly Pro Val Ser Gln Gly Lys Ala Ala Gly Gln  
 1890 1895 1900  
 Val Thr Pro Pro Thr Pro Gln Thr Ala Gln Pro Pro Leu Pro Gly  
 905 1910 1915 1920  
 Pro Pro Pro Thr Ala Val Glu Met Ala Met Gln Ile Gln Arg Ala Ala  
 1925 1930 1935  
 Glu Thr Gln Arg Gln Met Ala His Val Gln Ile Phe Gln Arg Pro Ile  
 1940 1945 1950  
 Gln His Gln Met Pro Pro Met Thr Pro Met Ala Pro Met Gly Met Asn  
 1955 1960 1965  
 Pro Pro Pro Met Thr Arg Gly Pro Ser Gly His Leu Glu Pro Gly Met  
 1970 1975 1980  
 Gly Pro Thr Gly Met Gln Gln Gln Pro Pro Trp Ser Gln Gly Gly Leu  
 985 1990 1995 2000  
 Pro Gln Pro Gln Gln Leu Gln Ser Gly Met Pro Arg Pro Ala Met Met  
 2005 2010 2015  
 Ser Val Ala Gln His Gly Gln Pro Leu Asn Met Ala Pro Gln Pro Gly  
 2020 2025 2030  
 Leu Gly Gln Val Gly Ile Ser Pro Leu Lys Pro Gly Thr Val Ser Gln  
 2035 2040 2045  
 Gln Ala Leu Gln Asn Leu Leu Arg Thr Leu Arg Ser Pro Ser Ser Pro  
 2050 2055 2060  
 Leu Gln Gln Gln Gln Val Leu Ser Ile Leu His Ala Asn Pro Gln Leu  
 065 2070 2075 2080  
 Leu Ala Ala Phe Ile Lys Gln Arg Ala Ala Lys Tyr Ala Asn Ser Asn  
 2085 2090 2095

Pro Gln Pro Ile Pro Gly Gln Pro Gly Met Pro Gln Gly Gln Pro Gly  
                   2100                                  2105                                  2110  
 Leu Gln Pro Pro Thr Met Pro Gly Gln Gln Gly Val His Ser Asn Pro  
                   2115                                  2120                                  2125  
 Ala Met Gln Asn Met Asn Pro Met Gln Ala Gly Val Gln Arg Ala Gly  
                   2130                                  2135                                  2140  
 Leu Pro Gln Gln Gln Pro Gln Gln Gln Leu Gln Pro Pro Met Gly Gly  
 145                                  2150                                  2155                                  2160  
 Met Ser Pro Gln Ala Gln Gln Met Asn Met Asn His Asn Thr Met Pro  
                   2165                                  2170                                  2175  
 Ser Gln Phe Arg Asp Ile Leu Arg Arg Gln Gln Met Met Gln Gln Gln  
                   2180                                  2185                                  2190  
 Gln Gln Gln Gly Ala Gly Pro Gly Ile Gly Pro Gly Met Ala Asn His  
                   2195                                  2200                                  2205  
 Asn Gln Phe Gln Gln Pro Gln Gly Val Gly Tyr Pro Pro Gln Pro Gln  
                   2210                                  2215                                  2220  
 Gln Arg Met Gln His His Met Gln Gln Met Gln Gln Gly Asn Met Gly  
 225                                  2230                                  2235                                  2240  
 Gln Ile Gly Gln Leu Pro Gln Ala Leu Gly Ala Glu Ala Gly Ala Ser  
                   2245                                  2250                                  2255  
 Leu Gln Ala Tyr Gln Gln Arg Leu Leu Gln Gln Gln Met Gly Ser Pro  
                   2260                                  2265                                  2270  
 Val Gln Pro Asn Pro Met Ser Pro Gln Gln His Met Leu Pro Asn Gln  
                   2275                                  2280                                  2285  
 Ala Gln Ser Pro His Leu Gln Gly Gln Gln Ile Pro Asn Ser Leu Ser  
                   2290                                  2295                                  2300  
 Asn Gln Val Arg Ser Pro Gln Pro Val Pro Ser Pro Arg Pro Gln Ser  
 305                                  2310                                  2315                                  2320  
 Gln Pro Pro His Ser Ser Pro Ser Pro Arg Met Gln Pro Gln Pro Ser  
                   2325                                  2330                                  2335  
 Pro His His Val Ser Pro Gln Thr Ser Ser Pro His Pro Gly Leu Val  
                   2340                                  2345                                  2350  
 Ala Ala Gln Ala Asn Pro Met Glu Gln Gly His Phe Ala Ser Pro Asp  
                   2355                                  2360                                  2365  
 Gln Asn Ser Met Leu Ser Gln Leu Ala Ser Asn Pro Gly Met Ala Asn  
                   2370                                  2375                                  2380  
 Leu His Gly Ala Ser Ala Thr Asp Leu Gly Leu Ser Thr Asp Asn Ser  
 385                                  2390                                  2395                                  2400  
 Asp Leu Asn Ser Asn Leu Ser Gln Ser Thr Leu Asp Ile His  
                   2405                                  2410                                  2

## (2) INFORMATION FOR SEQ ID NO:7:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2441 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: None

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Met Ala Glu Asn Leu Leu Asp Gly Pro Pro Asn Pro Lys Arg Ala Lys  
 1                                  5                                  10                                  15  
 Leu Ser Ser Pro Gly Phe Ser Ala Asn Asp Asn Thr Asp Phe Gly Ser  
                   20                                  25                                  30  
 Leu Phe Asp Leu Glu Asn Asp Leu Pro Asp Glu Leu Ile Pro Asn Gly  
                   35                                  40                                  45  
 Glu Leu Ser Leu Leu Asn Ser Gly Asn Leu Val Pro Asp Ala Ala Ser  
 50                                  55                                  60  
 Lys His Lys Gln Leu Ser Glu Leu Leu Arg Gly Gly Ser Gly Ser Ser  
 65                                  70                                  75                                  80

Ile Asn Pro Gly Ile Gly Asn Val Ser Ala Ser Ser Pro Val Gln Gln  
 85 90 95  
 Gly Leu Gly Gly Gln Ala Gln Gly Gln Pro Asn Ser Thr Asn Met Ala  
 100 105 110  
 Ser Leu Gly Ala Met Gly Lys Ser Pro Leu Asn Gln Gly Asp Ser Ser  
 115 120 125  
 Thr Pro Asn Leu Pro Lys Gln Ala Ala Ser Thr Ser Gly Pro Thr Pro  
 130 135 140  
 Pro Ala Ser Gln Ala Leu Asn Pro Gln Ala Gln Lys Gln Val Gly Leu  
 145 150 155 160  
 Val Thr Ser Ser Pro Ala Thr Ser Gln Thr Gly Pro Gly Ile Cys Met  
 165 170 175  
 Asn Ala Asn Phe Asn Gln Thr His Pro Gly Leu Leu Asn Ser Asn Ser  
 180 185 190  
 Gly His Ser Leu Met Asn Gln Ala Gln Gln Gly Gln Ala Gln Val Met  
 195 200 205  
 Asn Gly Ser Leu Gly Ala Ala Gly Arg Gly Arg Gly Ala Gly Met Pro  
 210 215 220  
 Tyr Pro Ala Pro Ala Met Gln Gly Ala Thr Ser Ser Val Leu Ala Glu  
 225 230 235 240  
 Thr Leu Thr Gln Val Ser Pro Gln Met Ala Gly His Ala Gly Leu Asn  
 245 250 255  
 Thr Ala Gln Ala Gly Gly Met Thr Lys Met Gly Met Thr Gly Thr Thr  
 260 265 270  
 Ser Pro Phe Gly Gln Pro Phe Ser Gln Thr Gly Gly Gln Gln Met Gly  
 275 280 285  
 Ala Thr Gly Val Asn Pro Gln Leu Ala Ser Lys Gln Ser Met Val Asn  
 290 295 300  
 Ser Leu Pro Ala Phe Pro Thr Asp Ile Lys Asn Thr Ser Val Thr Thr  
 305 310 315 320  
 Val Pro Asn Met Ser Gln Leu Gln Thr Ser Val Gly Ile Val Pro Thr  
 325 330 335  
 Gln Ala Ile Ala Thr Gly Pro Thr Ala Asp Pro Glu Lys Arg Lys Leu  
 340 345 350  
 Ile Gln Gln Gln Leu Val Leu Leu Leu His Ala His Lys Cys Gln Arg  
 355 360 365  
 Arg Glu Gln Ala Asn Gly Glu Val Arg Ala Cys Ser Leu Pro His Cys  
 370 375 380  
 Arg Thr Met Lys Asn Val Leu Asn His Met Thr His Cys Gln Ala Pro  
 385 390 395 400  
 Lys Ala Cys Gln Val Ala His Cys Ala Ser Ser Arg Gln Ile Ile Ser  
 405 410 415  
 His Trp Lys Asn Cys Thr Arg His Asp Cys Pro Val Cys Leu Pro Leu  
 420 425 430  
 Lys Asn Ala Ser Asp Lys Arg Asn Gln Gln Thr Ile Leu Gly Ser Pro  
 435 440 445  
 Ala Ser Gly Ile Gln Asn Thr Ile Gly Ser Val Gly Ala Gly Gln Gln  
 450 455 460  
 Asn Ala Thr Ser Leu Ser Asn Pro Asn Pro Ile Asp Pro Ser Ser Met  
 465 470 475 480  
 Gln Arg Ala Tyr Ala Ala Leu Gly Leu Pro Tyr Met Asn Gln Pro Gln  
 485 490 495  
 Thr Gln Leu Gln Pro Gln Val Pro Gly Gln Gln Pro Ala Gln Pro Pro  
 500 505 510  
 Ala His Gln Gln Met Arg Thr Leu Asn Ala Leu Gly Asn Asn Pro Met  
 515 520 525  
 Ser Val Pro Ala Gly Gly Ile Thr Thr Asp Gln Gln Pro Pro Asn Leu  
 530 535 540  
 Ile Ser Glu Ser Ala Leu Pro Thr Ser Leu Gly Ala Thr Asn Pro Leu  
 545 550 555 560  
 Met Asn Asp Gly Ser Asn Ser Gly Asn Ile Gly Ser Leu Ser Thr Ile  
 565 570 575

Pro	Thr	Ala	Ala	Pro	Pro	Ser	Ser	Thr	Gly	Val	Arg	Lys	Gly	Trp	His
			580						585				590		
Glu	His	Val	Thr	Gln	Asp	Leu	Arg	Ser	His	Leu	Val	His	Lys	Leu	Val
		595					600					605			
Gln	Ala	Ile	Phe	Pro	Thr	Pro	Asp	Pro	Ala	Ala	Leu	Lys	Asp	Arg	Arg
	610					615					620				
Met	Glu	Asn	Leu	Val	Ala	Tyr	Ala	Lys	Lys	Val	Glu	Gly	Asp	Met	Tyr
625					630					635					640
Glu	Ser	Ala	Asn	Ser	Arg	Asp	Glu	Tyr	Tyr	His	Leu	Leu	Ala	Glu	Lys
			645						650					655	
Ile	Tyr	Lys	Ile	Gln	Lys	Glu	Leu	Glu	Glu	Lys	Arg	Arg	Thr	Arg	Leu
			660					665					670		
His	Lys	Gln	Gly	Ile	Leu	Gly	Asn	Gln	Pro	Ala	Leu	Pro	Ala	Ser	Gly
		675					680					685			
Ala	Gln	Pro	Pro	Val	Ile	Pro	Pro	Ala	Gln	Ser	Val	Arg	Pro	Pro	Asn
		690				695					700				
Gly	Pro	Leu	Pro	Leu	Pro	Val	Asn	Arg	Met	Gln	Val	Ser	Gln	Gly	Met
705					710					715					720
Asn	Ser	Phe	Asn	Pro	Met	Ser	Leu	Gly	Asn	Val	Gln	Leu	Pro	Gln	Ala
			725						730					735	
Pro	Met	Gly	Pro	Arg	Ala	Ala	Ser	Pro	Met	Asn	His	Ser	Val	Gln	Met
			740					745					750		
Asn	Ser	Met	Ala	Ser	Val	Pro	Gly	Met	Ala	Ile	Ser	Pro	Ser	Arg	Met
		755					760					765			
Pro	Gln	Pro	Pro	Asn	Met	Met	Gly	Thr	His	Ala	Asn	Asn	Ile	Met	Ala
		770				775					780				
Gln	Ala	Pro	Thr	Gln	Asn	Gln	Phe	Leu	Pro	Gln	Asn	Gln	Phe	Pro	Ser
785					790					795					800
Ser	Ser	Gly	Ala	Met	Ser	Val	Asn	Ser	Val	Gly	Met	Gly	Gln	Pro	Ala
			805						810					815	
Ala	Gln	Ala	Gly	Val	Ser	Gln	Gly	Gln	Glu	Pro	Gly	Ala	Ala	Leu	Pro
			820					825					830		
Asn	Pro	Leu	Asn	Met	Leu	Ala	Pro	Gln	Ala	Ser	Gln	Leu	Pro	Cys	Pro
		835					840					845			
Pro	Val	Thr	Gln	Ser	Pro	Leu	His	Pro	Thr	Pro	Pro	Pro	Ala	Ser	Thr
		850				855					860				
Ala	Ala	Gly	Met	Pro	Ser	Leu	Gln	His	Pro	Thr	Ala	Pro	Gly	Met	Thr
865					870					875					880
Pro	Pro	Gln	Pro	Ala	Ala	Pro	Thr	Gln	Pro	Ser	Thr	Pro	Val	Ser	Ser
			885						890					895	
Gly	Gln	Thr	Pro	Thr	Pro	Thr	Pro	Gly	Ser	Val	Pro	Ser	Ala	Ala	Gln
			900					905					910		
Thr	Gln	Ser	Thr	Pro	Thr	Val</									

Ser Gln Ser Thr Ser Pro Ser Gln Pro Arg Lys Lys Ile Phe Lys Pro  
 1075 1080 1085  
 Glu Glu Leu Arg Gln Ala Leu Met Pro Thr Leu Glu Ala Leu Tyr Arg  
 1090 1095 1100  
 Gln Asp Pro Glu Ser Leu Pro Phe Arg Gln Pro Val Asp Pro Gln Leu  
 105 1110 1115 1120  
 Leu Gly Ile Pro Asp Tyr Phe Asp Ile Val Lys Asn Pro Met Asp Leu  
 1125 1130 1135  
 Ser Thr Ile Lys Arg Lys Leu Asp Thr Gly Gln Tyr Gln Glu Pro Trp  
 1140 1145 1150  
 Gln Tyr Val Asp Asp Val Arg Leu Met Phe Asn Asn Ala Trp Leu Tyr  
 1155 1160 1165  
 Asn Arg Lys Thr Ser Arg Val Tyr Lys Phe Cys Ser Lys Leu Ala Glu  
 1170 1175 1180  
 Val Phe Glu Gln Glu Ile Asp Pro Val Met Gln Ser Leu Gly Tyr Cys  
 185 1190 1195 1200  
 Cys Gly Arg Lys Tyr Glu Phe Ser Pro Gln Thr Leu Cys Cys Tyr Gly  
 1205 1210 1215  
 Lys Gln Leu Cys Thr Ile Pro Arg Asp Ala Ala Tyr Tyr Ser Tyr Gln  
 1220 1225 1230  
 Asn Arg Tyr His Phe Cys Gly Lys Cys Phe Thr Glu Ile Gln Gly Glu  
 1235 1240 1245  
 Asn Val Thr Leu Gly Asp Asp Pro Ser Gln Pro Gln Thr Thr Ile Ser  
 1250 1255 1260  
 Lys Asp Gln Phe Glu Lys Lys Lys Asn Asp Thr Leu Asp Pro Glu Pro  
 265 1270 1275 1280  
 Phe Val Asp Cys Lys Glu Cys Gly Arg Lys Met His Gln Ile Cys Val  
 1285 1290 1295  
 Leu His Tyr Asp Ile Ile Trp Pro Ser Gly Phe Val Cys Asp Asn Cys  
 1300 1305 1310  
 Leu Lys Lys Thr Gly Arg Pro Arg Lys Glu Asn Lys Phe Ser Ala Lys  
 1315 1320 1325  
 Arg Leu Gln Thr Thr Arg Leu Gly Asn His Leu Glu Asp Arg Val Asn  
 1330 1335 1340  
 Lys Phe Leu Arg Arg Gln Asn His Pro Glu Ala Gly Glu Val Phe Val  
 345 1350 1355 1360  
 Arg Val Val Ala Ser Ser Asp Lys Thr Val Glu Val Lys Pro Gly Met  
 1365 1370 1375  
 Lys Ser Arg Phe Val Asp Ser Gly Glu Met Ser Glu Ser Phe Pro Tyr  
 1380 1385 1390  
 Arg Thr Lys Ala Leu Phe Ala Phe Glu Glu Ile Asp Gly Val Asp Val  
 1395 1400 1405  
 Cys Phe Phe Gly Met His Val Gln Asp Thr Ala Leu Ile Ala Pro His  
 1410 1415 1420  
 Gln Ile Gln Gly Cys Val Tyr Ile Ser Tyr Leu Asp Ser Ile His Phe  
 425 1430 1435 1440  
 Phe Arg Pro Arg Cys Leu Arg Thr Ala Val Tyr His Glu Ile Leu Ile  
 1445 1450 1455  
 Gly Tyr Leu Glu Tyr Val Lys Lys Leu Val Tyr Val Thr Ala His Ile  
 1460 1465 1470  
 Trp Ala Cys Pro Pro Ser Glu Gly Asp Asp Tyr Ile Phe His Cys His  
 1475 1480 1485  
 Pro Pro Asp Gln Lys Ile Pro Lys Pro Lys Arg Leu Gln Glu Trp Tyr  
 1490 1495 1500  
 Lys Lys Met Leu Asp Lys Ala Phe Ala Glu Arg Ile Ile Asn Asp Tyr  
 505 1510 1515 1520  
 Lys Asp Ile Phe Lys Gln Ala Asn Glu Asp Arg Leu Thr Ser Ala Lys  
 1525 1530 1535  
 Glu Leu Pro Tyr Phe Glu Gly Asp Phe Trp Pro Asn Val Leu Glu Glu  
 1540 1545 1550  
 Ser Ile Lys Glu Leu Glu Gln Glu Glu Glu Arg Lys Lys Glu Glu  
 1555 1560 1565

Ser Thr Ala Ala Ser Glu Thr Pro Glu Gly Ser Gln Gly Asp Ser Lys  
 1570 1575 1580  
 Asn Ala Lys Lys Lys Asn Asn Lys Lys Thr Asn Lys Asn Lys Ser Ser  
 585 1590 1595 1600  
 Ile Ser Arg Ala Asn Lys Lys Lys Pro Ser Met Pro Asn Val Ser Asn  
 1605 1610 1615  
 Asp Leu Ser Gln Lys Leu Tyr Ala Thr Met Glu Lys His Lys Glu Val  
 1620 1625 1630  
 Phe Phe Val Ile His Leu His Ala Gly Pro Val Ile Ser Thr Gln Pro  
 1635 1640 1645  
 Pro Ile Val Asp Pro Asp Pro Leu Leu Ser Cys Asp Leu Met Asp Gly  
 1650 1655 1660  
 Arg Asp Ala Phe Leu Thr Leu Ala Arg Asp Lys His Trp Glu Phe Ser  
 665 1670 1675 1680  
 Ser Leu Arg Arg Ser Lys Trp Ser Thr Leu Cys Met Leu Val Glu Leu  
 1685 1690 1695  
 His Thr Gln Gly Gln Asp Arg Phe Val Tyr Thr Cys Asn Glu Cys Lys  
 1700 1705 1710  
 His His Val Glu Thr Arg Trp His Cys Thr Val Cys Glu Asp Tyr Asp  
 1715 1720 1725  
 Leu Cys Ile Asn Cys Tyr Asn Thr Lys Ser His Thr His Lys Met Val  
 1730 1735 1740  
 Lys Trp Gly Leu Gly Leu Asp Asp Glu Gly Ser Ser Gln Gly Glu Pro  
 745 1750 1755 1760  
 Gln Ser Lys Ser Pro Gln Glu Ser Arg Arg Leu Ser Ile Gln Arg Cys  
 1765 1770 1775  
 Ile Gln Ser Leu Val His Ala Cys Gln Cys Arg Asn Ala Asn Cys Ser  
 1780 1785 1790  
 Leu Pro Ser Cys Gln Lys Met Lys Arg Val Val Gln His Thr Lys Gly  
 1795 1800 1805  
 Cys Lys Arg Lys Thr Asn Gly Gly Cys Pro Val Cys Lys Gln Leu Ile  
 1810 1815 1820  
 Ala Leu Cys Cys Tyr His Ala Lys His Cys Gln Glu Asn Lys Cys Pro  
 825 1830 1835 1840  
 Val Pro Phe Cys Leu Asn Ile Lys His Asn Val Arg Gln Gln Gln Ile  
 1845 1850 1855  
 Gln His Cys Leu Gln Gln Ala Gln Leu Met Arg Arg Arg Met Ala Thr  
 1860 1865 1870  
 Met Asn Thr Arg Asn Val Pro Gln Gln Ser Leu Pro Ser Pro Thr Ser  
 1875 1880 1885  
 Ala Pro Pro Gly Thr Pro Thr Gln Gln Pro Ser Thr Pro Gln Thr Pro  
 1890 1895 1900  
 Gln Pro Pro Ala Gln Pro Gln Pro Ser Pro Val Asn Met Ser Pro Ala  
 905 1910 1915 1920  
 Gly Phe Pro Asn Val Ala Arg Thr Gln Pro Pro Thr Ile Val Ser Ala  
 1925 1930 1935  
 Gly Lys Pro Thr Asn Gln Val Pro Ala Pro Pro Pro Ala Gln Pro  
 1940 1945 1950  
 Pro Pro Ala Ala Val Glu Ala Ala Arg Gln Ile Glu Arg Glu Ala Gln  
 1955 1960 1965  
 Gln Gln Gln His Leu Tyr Arg Ala Asn Ile Asn Asn Gly Met Pro Pro  
 1970 1975 1980  
 Gly Arg Asp Gly Met Gly Thr Pro Gly Ser Gln Met Thr Pro Val Gly  
 985 1990 1995 2000  
 Leu Asn Val Pro Arg Pro Asn Gln Val Ser Gly Pro Val Met Ser Ser  
 2005 2010 2015  
 Met Pro Pro Gly Gln Trp Gln Gln Ala Pro Ile Pro Gln Gln Gln Pro  
 2020 2025 2030  
 Met Pro Gly Met Pro Arg Pro Val Met Ser Met Gln Ala Gln Ala Ala  
 2035 2040 2045  
 Val Ala Gly Pro Arg Met Pro Asn Val Gln Pro Asn Arg Ser Ile Ser  
 2050 2055 2060

Pro Ser Ala Leu Gln Asp Leu Leu Arg Thr Leu Lys Ser Pro Ser Ser  
 065 2070 2075 2080  
 Pro Gln Gln Gln Gln Val Leu Asn Ile Leu Lys Ser Asn Pro Gln  
 2085 2090 2095  
 Leu Met Ala Ala Phe Ile Lys Gln Arg Thr Ala Lys Tyr Val Ala Asn  
 2100 2105 2110  
 Gln Pro Gly Met Gln Pro Gln Pro Gly Leu Gln Ser Gln Pro Gly Met  
 2115 2120 2125  
 Gln Pro Gln Pro Gly Met His Gln Gln Pro Ser Leu Gln Asn Leu Asn  
 2130 2135 2140  
 Ala Met Gln Ala Gly Val Pro Arg Pro Gly Val Pro Pro Pro Gln Pro  
 145 2150 2155 2160  
 Ala Met Gly Gly Leu Asn Pro Gln Gly Gln Ala Leu Asn Ile Met Asn  
 2165 2170 2175  
 Pro Gly His Asn Pro Asn Met Thr Asn Met Asn Pro Gln Tyr Arg Glu  
 2180 2185 2190  
 Met Val Arg Arg Gln Leu Leu Gln His Gln Gln Gln Gln Gln Gln  
 2195 2200 2205  
 Gln Gln Gln Gln Gln Gln Gln Gln Asn Ser Ala Ser Leu Ala Gly Gly  
 2210 2215 2220  
 Met Ala Gly His Ser Gln Phe Gln Gln Pro Gln Gly Pro Gly Gly Tyr  
 225 2230 2235 2240  
 Ala Pro Ala Met Gln Gln Gln Arg Met Gln Gln His Leu Pro Ile Gln  
 2245 2250 2255  
 Gly Ser Ser Met Gly Gln Met Ala Ala Pro Met Gly Gln Leu Gly Gln  
 2260 2265 2270  
 Met Gly Gln Pro Gly Leu Gly Ala Asp Ser Thr Pro Asn Ile Gln Gln  
 2275 2280 2285  
 Ala Leu Gln Gln Arg Ile Leu Gln Gln Gln Gln Met Lys Gln Gln Ile  
 2290 2295 2300  
 Gly Ser Pro Gly Gln Pro Asn Pro Met Ser Pro Gln Gln His Met Leu  
 305 2310 2315 2320  
 Ser Gly Gln Pro Gln Ala Ser His Leu Pro Gly Gln Gln Ile Ala Thr  
 2325 2330 2335  
 Ser Leu Ser Asn Gln Val Arg Ser Pro Ala Pro Val Gln Ser Pro Arg  
 2340 2345 2350  
 Pro Gln Ser Gln Pro Pro His Ser Ser Pro Ser Pro Arg Ile Gln Pro  
 2355 2360 2365  
 Gln Pro Ser Pro His His Val Ser Pro Gln Thr Gly Thr Pro His Pro  
 2370 2375 2380  
 Gly Leu Ala Val Thr Met Ala Ser Ser Met Asp Gln Gly His Leu Gly  
 385 2390 2395 2400  
 Asn Pro Glu Gln Ser Ala Met Leu Pro Gln Leu Asn Thr Pro Asn Arg  
 2405 2410 2415  
 Ser Ala Leu Ser Ser Glu Leu Ser Leu Val Gly Asp Thr Thr Gly Asp  
 2420 2425 2430  
 Thr Leu Glu Lys Phe Val Glu Gly Leu  
 2435 2440

## (2) INFORMATION FOR SEQ ID NO:8:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 813 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: None

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Met Ala Glu Ala Gly Gly Ala Gly Ser Pro Ala Leu Pro Pro Ala Pro  
 1 5 10 15

Pro	His	Gly	Ser	Pro	Arg	Thr	Leu	Ala	Thr	Ala	Ala	Gly	Ser	Ser	Ala		
			20					25					30				
Ser	Cys	Gly	Pro	Ala	Thr	Pro	Val	Ala	Ala	Ala	Gly	Thr	Ala	Glu	Gly		
		35					40					45					
Pro	Gly	Gly	Gly	Gly	Ser	Ala	Arg	Ile	Ala	Val	Lys	Lys	Ala	Gln	Leu		
		50				55					60						
Arg	Ser	Ala	Pro	Arg	Ala	Lys	Lys	Leu	Glu	Lys	Leu	Gly	Val	Tyr	Ser		
		65				70				75					80		
Ala	Cys	Lys	Ala	Glu	Glu	Ser	Cys	Lys	Cys	Asn	Gly	Trp	Lys	Asn	Pro		
			85						90					95			
Asn	Pro	Ser	Pro	Thr	Pro	Pro	Arg	Gly	Asp	Leu	Gln	Gln	Ile	Ile	Val		
			100					105					110				
Ser	Leu	Thr	Glu	Ser	Cys	Arg	Ser	Cys	Ser	His	Ala	Leu	Ala	Ala	His		
		115						120				125					
Val	Ser	His	Leu	Glu	Asn	Val	Ser	Glu	Glu	Glu	Met	Asp	Arg	Leu	Leu		
		130				135					140						
Gly	Ile	Val	Leu	Asp	Val	Glu	Tyr	Leu	Phe	Thr	Cys	Val	His	Lys	Glu		
		145				150				155					160		
Glu	Asp	Ala	Asp	Thr	Lys	Gln	Val	Tyr	Phe	Tyr	Leu	Phe	Lys	Leu	Leu		
			165						170					175			
Arg	Lys	Ser	Ile	Leu	Gln	Arg	Gly	Lys	Pro	Val	Val	Glu	Gly	Ser	Leu		
			180					185						190			
Glu	Lys	Lys	Pro	Pro	Phe	Glu	Lys	Pro	Ser	Ile	Glu	Gln	Gly	Val	Asn		
		195					200					205					
Asn	Phe	Val	Gln	Tyr	Lys	Phe	Ser	His	Leu	Pro	Ser	Lys	Glu	Arg	Gln		
		210				215					220						
Thr	Thr	Ile	Glu	Leu	Ala	Lys	Met	Phe	Leu	Asn	Arg	Ile	Asn	Tyr	Trp		
					230					235					240		
His	Leu	Glu	Ala	Pro	Ser	Gln	Arg	Arg	Leu	Arg	Ser	Pro	Asn	Asp	Asp		
			245					250						255			
Ile	Ser	Gly	Tyr	Lys	Glu	Asn	Tyr	Thr	Arg	Trp	Leu	Cys	Tyr	Cys	Asn		
			260				265						270				
Val	Pro	Gln	Phe	Cys	Asp	Ser	Leu	Pro	Arg	Tyr	Glu	Thr	Thr	Lys	Val		
		275					280					285					
Phe	Gly	Arg	Thr	Leu	Leu	Arg	Ser	Val	Phe	Thr	Ile	Met	Arg	Arg	Gln		
		290				295					300						
Leu	Leu	Glu	Gln	Ala	Arg	Gln	Lys	Lys	Asp	Lys	Leu	Pro	Leu	Glu	Lys		
					310					315					320		
Arg	Thr	Leu	Ile	Leu	Thr	His	Phe	Pro	Lys	Phe	Leu	Ser	Met	Leu	Glu		
				325					330					335			
Glu	Glu	Val	Tyr	Ser	Gln	Asn	Ser	Pro	Ile	Trp	Asp	Gln	Asp	Phe	Leu		
			340					345					350				
Ser	Ala	Ser	Ser	Arg	Thr	Ser	Pro	Leu	Gly	Ile	Gln	Thr	Val	Ile	Ser		
		355					360					365					
Pro	Pro	Val	Thr	Gly	Thr	Ala	Leu	Phe	Ser	Ser	Asn	Ser	Thr	Ser	His		
						375					380						
Glu	Gln	Ile	Asn	Gly	Gly	Arg	Thr	Ser	Pro	Gly	Cys	Arg	Gly	Ser	Ser		
				390						395					400		
Gly	Leu	Glu	Ala	Asn	Pro	Gly	Glu	Lys	Arg	Lys	Met	Asn	Asn	Ser	His		
				405					410					415			
Ala	Pro	Glu	Glu	Ala	Lys	Arg	Ser	Arg	Val	Met	Gly	Asp	Ile	Pro	Val		
				420				425					430				
Glu	Leu	Ile	Asn	Glu	Val	Met	Ser	Thr	Ile	Thr	Asp	Pro	Ala	Gly	Met		
		435					440					445					
Leu	Gly	Pro	Glu	Thr	Asn	Phe	Leu	Ser	Ala	His	Ser	Ala	Arg	Asp	Glu		
		450				455					460						
Ala	Ala	Arg	Leu	Glu	Glu	Arg	Arg	Gly	Val	Ile	Glu	Phe	His	Val	Val		
					470					475					480		
Gly	Asn	Ser	Leu	Asn	Gln	Lys	Pro	Asn	Lys	Lys	Ile	Leu	Met	Trp	Leu		
				485					490					495			
Val	Gly	Leu	Gln	Asn	Val	Phe	Ser	His	Gln	Leu	Pro	Arg	Met	Pro	Lys		
			500					505					510				

Glu Tyr Ile Thr Arg Leu Val Phe Asp Pro Lys His Lys Thr Leu Ala  
           515                          520                          525  
 Leu Ile Lys Asp Gly Arg Val Ile Gly Gly Ile Cys Phe Arg Met Phe  
       530                          535                          540  
 Pro Ser Gln Gly Phe Thr Glu Ile Val Phe Cys Ala Val Thr Ser Asn  
 545                          550                          555                          560  
 Glu Gln Val Lys Gly Tyr Gly Thr His Leu Met Asn His Leu Lys Glu  
                           565                          570                          575  
 Tyr His Ile Lys His Glu Ile Leu Asn Phe Leu Thr Tyr Ala Asp Glu  
                           580                          585                          590  
 Tyr Ala Ile Gly Tyr Phe Lys Lys Gln Gly Phe Ser Lys Glu Ile Lys  
                           595                          600                          605  
 Ile Pro Lys Thr Lys Tyr Val Gly Tyr Ile Lys Asp Tyr Glu Gly Ala  
       610                          615                          620  
 Thr Leu Met Gly Cys Glu Leu Asn Pro Gln Ile Pro Tyr Thr Glu Phe  
 625                          630                          635                          640  
 Ser Val Ile Ile Lys Lys Gln Lys Glu Ile Lys Lys Leu Ile Glu  
                           645                          650                          655  
 Arg Lys Gln Ala Gln Ile Arg Lys Val Tyr Pro Gly Leu Ser Cys Phe  
                           660                          665                          670  
 Lys Asp Gly Val Arg Gln Ile Pro Ile Glu Ser Ile Pro Gly Ile Arg  
                           675                          680                          685  
 Glu Thr Gly Trp Lys Pro Ser Gly Lys Glu Lys Ser Lys Glu Pro Lys  
       690                          695                          700  
 Asp Pro Glu His Val Tyr Ser Thr Leu Lys Asn Ile Leu Gln Gln Val  
 705                          710                          715                          720  
 Lys Asn His Pro Asn Ala Trp Pro Phe Met Glu Pro Val Lys Arg Thr  
                           725                          730                          735  
 Glu Ala Pro Gly Tyr Tyr Glu Val Ile Arg Phe Pro Met Asp Leu Lys  
                           740                          745                          750  
 Thr Met Ser Glu Arg Leu Arg Asn Arg Tyr Tyr Val Ser Lys Lys Leu  
                           755                          760                          765  
 Phe Met Ala Asp Leu Gln Arg Val Phe Thr Asn Cys Lys Glu Tyr Asn  
       770                          775                          780  
 Pro Pro Glu Ser Glu Tyr Tyr Lys Cys Ala Ser Ile Leu Glu Lys Phe  
 785                          790                          795                          800  
 Phe Phe Ser Lys Ile Lys Glu Ala Gly Leu Ile Asp Lys  
                           805                          810

## (2). INFORMATION FOR SEQ ID NO:9:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

His Thr Lys Gly Cys Lys Arg Lys Thr Asn Gly Gly Cys Pro Val Cys  
   1                          5                          10                          15  
 Lys Gln Leu Ile Ala Leu Cys Cys Tyr His Ala Lys His Cys Gln Glu  
                           20                          25                          30  
 Asn Lys Cys Pro Val Pro Phe Cys Leu Asn Ile Lys His Asn Val Arg  
                           35                          40                          45  
 Gln Gln  
   50

## (2) INFORMATION FOR SEQ ID NO:10:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2204 base pairs

- (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

ACCCACTCCC	CCCAGAGCCG	ACCTGCAGCA	AATAATTGTC	AGTCTAACAG	AATCCTGTCTG	60
GAGTTGTAGC	CATGCCCTAG	CTGCTCATGT	TTCCCACCTG	GAGAATGTGT	CAGAGGAAGA	120
AATGAACAGA	CTCCTGGGAA	TAGTATTGGA	TGTGGAATAT	CTCTTTACCT	GTGTCCACAA	180
GGAAGAAGAT	GCAGATACCA	AACAAGTTTA	TTTCTATCTA	TTTAAGCTCT	TGAGAAAGTC	240
TATTTTACAA	AGAGGAAAAC	CTGTGGTTGG	AAGGCTCTTT	GGAAAAGAAA	CCCCCATTTG	300
AAAAACCTAG	CATTGAACAG	GGTGTGAATA	ACTTTGTGCA	GTACAAATTT	AGTCACCTGC	360
CAGCAAAAAG	AAAGGCAAAC	CAATAGTTGA	GTTGGCAAAA	ATGTTCTTAA	ACCGCATCAC	420
CTATTGGCAT	CTGGAGGCAC	CATCTCAACG	AGACTGCGAT	CTCCAATGAT	GATATTCTGG	480
ATACAAAGAG	AACTACACAA	GGTGGCTGTG	TTACTGCAAC	GTGCCACAGT	TCTGCGACAG	540
TCTACCTCGG	TACGAAACCA	CACAGGTGTT	TGGGAGAACA	TCGTTGCTC	GGTCTTCACT	600
GTTATGAGGC	GACAACTCCT	GGAACAAGCA	AGACAGGAAA	AAGATAAACT	GCCTCTTGAA	660
AAACGAACCT	TATTCCTCAC	TCATTTCCCA	AAATTTCTGT	CCATGCTAGA	AGAAGAAGTA	720
TATAGTCAAA	ACTCTCCCAT	CTGGGATCAC	CATTTTCTCT	CAGCCTCTTC	CAGAACCAGC	780
CAGCTAGGCA	TCCAAACAGT	TATCAATCAC	CTCCTGTGGC	TGGGACAATT	TCATACAATT	840
CAACCTCATC	TTCCCTTGAG	CAGCCAAACG	CAGGGAGCAG	CAGTCCTGCC	TGCAAAGCCT	900
CTTCTGGACT	TGAGGCAAAC	CCAGGAGAAA	AGAGGAAAAT	GACTGATTCT	CATGTTCTGG	960
AGGAGGCCAA	GAAACCCCGA	GTTATGGGGG	ATATTCGGAT	GGAATTAATC	AACGAGGTTA	1020
TGCTTACCAT	CACGGACCCCT	GCAGCAATGC	TTGGACCAGA	GACCAATTTT	TCTGCAGCAC	1080
ACTCGGCCAG	GGATGAGGCG	GCAAGGTTGG	AAGAGCGCAG	GGGTGTAATT	GAATTTACAG	1140
TGGTTGGCAA	TTCCCTCAAC	CAGAAACCAA	ACAAGAAGAT	CCTGATGTGG	CTGGTTGGCC	1200
TACAGAACGT	TTTCTCCAC	CAGCTGCCCC	GAATGCCAAA	AGAATACATC	ACACGGCTCG	1260
TCTTTGACCC	GAAACACAAA	ACCCTTGCTT	TAATTTAAAG	TGGCCGTGTT	ATTGGTGGTA	1320
TCTGTTTCCG	TATGTTCCCA	TCTCAAGGAT	TCACAGAGAT	TGTCTTCTGT	GCTGTAACCT	1380
CAAATGAGCA	AGTCAAGGCG	TATGGAACAC	ACCTGATGAA	TCATTTGAAA	GAATATCACA	1440
TAAAGCATGA	CATCCTGAAC	TTCTCAGCAT	ATGCAGATGA	ATATGCAATT	GGATACTTTA	1500
AGAAACAGGG	TTTCTCCAAA	GAAATTAATA	TACCTAAAC	CAAATATGTT	GGCTATATCA	1560
AGGATTATGA	AGGAGCCACT	TTAATGGGAT	GTGAGCTAAA	TCCACGGATC	CCGTACACAG	1620
AATTTTCTGT	CATCATTAAT	AAGCAGAAGG	AGATAATTAA	AAAACCTGAT	GAAAGAAAAC	1680
AGGCACAAAT	TCGAAAAGTT	TACCCTGGAC	TTTCATGTTT	TAAAGATGGA	GTTCGACAGA	1740
TTCTTATAGA	AAGCATTCCCT	GGAATTAGAG	AGACAGGCTG	GAAACCGAGT	GGAAAAGAGA	1800
AAAGTAAAGA	GCCCAGAGAC	CCTGACCAGC	TTTACAGCAC	GCTCAAGAGC	ATCCTCCAGC	1860
AGGTGAAGAG	CCATCAAAGC	GCTTGGCCCT	TCATGGAACC	TGTGAAGAGA	ACAGAAGCTC	1920
CAGGATATTA	TGAAGTTATA	AGGTCCCCCA	TGGATCTCAA	AACCATGAGT	GAACGCCTCA	1980
AGAATAGGTA	CTACGTGTCT	AAGAAATTAT	TCATGGCAGA	CTTACAGCGA	GTCTTTACCA	2040
ATTGCAAAGA	GTACAACGCC	CCTGAGAGTG	AATACTACAA	ATGTGCCAAT	ATCCTGGAGA	2100
AATCTTCTTT	CAGTAAAATT	AAGGAAGCTG	GATTAATTGA	CAAGTGATTT	TTTTTCCCCC	2160
TCTGCTTCTT	AGAAACTCAC	CAAGCAGTGT	GCCTAAAGCA	AGGT		2204

## (2) INFORMATION FOR SEQ ID NO:11:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2093 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GAATTCGGGC	GAAACCACTC	ATGTCTTTGG	GCGAAGCCTT	CTCCGGTCCA	TTTTACCCGT	60
TACCCGCGCG	CAGCTGCTGG	AAAAGTTCCG	AGTGGAGAAG	GACAAATTGG	TGCCCCGAGAA	120
GAGGACCCCT	ATCCTCACTC	ACTTCCCCAA	GTAAGGCTCC	TTCTGGCCTA	CCAGGATTTG	180
GCCCCAAGTT	CACATCCTCC	CTGTTGTCCC	CTTTTCTCCA	GGAAGGCTTC	CTGGATTGGT	240
CCCTCCTCTC	CCTCCATGGG	CCTTTTGGGA	TCTGGGCGTC	TACCTGGCAG	ACTTGCCCAT	300
GGCCCAGAAG	CAACTTGCTA	GTACTAGTCT	GGGGATGGCA	GATTCTGTCT	CATGCTGGAG	360
GAGGAGATCT	ATGGGGCAAA	CTCTCCAATC	TGGGAGTCAG	GCTTCACCAT	GCCACCCTCA	420

GAGGGGACAC	AGCTGGTTCC	CCGGCCAGCT	TCAGTCAGTG	CAGCGGTTGT	TCCCAGCACC	480
CCCATCTTCA	GCCCCAGCAT	GGGTGGGGGC	AGCAACAGCT	CCCTGAGTCT	GGATTCTGCA	540
GGGGCCGAGC	CTATGCCAGG	CGAGAAGAGG	ACGCTCCCAG	AGAACCTGAC	CCTGGAGGAT	600
GCCAAGCGGC	TCCGTGTGAT	GGGTGACATC	CCCATGGAGC	TGGTCAATGA	GGTCATGCTG	660
ACCATCACTG	ACCCTGCTGC	CATGCTGGGG	CCTGAGACGA	GCCTGCTTTC	GGCCAATGCG	720
GCCCGGGATG	AGACAGCCCG	CCTGGAGGAG	CGCCGCGGCA	TCATCGAGTT	CCATGTCATC	780
GGCAACTCAC	TGACGCCCAA	GGCCAACCGG	CGGGTGTTCG	TGTGGCTCGT	GGGGCTGCAG	840
AATGTCTTTT	CCCACCAGCT	GCCGCGCATG	CCTAAGGAGT	ATATCGCCCC	CCTCGTCTTT	900
GACCCGAAGC	ACAAGACTCT	GGCCTTGATC	AAGGATGGGC	GGGTCATCGG	TGGCATCTGC	960
TTCCGCGATG	TTCCACCCCA	GGGCTTCACG	GAGATTGTCT	TCTGTGCTGT	CACCTCGAAT	1020
GAGCAGGTCA	AGGGTTATGG	GACCCACCTG	ATGAACCACC	TGAAGGAGTA	TCACATCAAG	1080
CACAACATTC	TCTACTTCCT	CACCTACGCC	GACGAGTACG	CCATCGGCTA	CTTCAAAAAG	1140
CAGGGTTTCT	CCAAGGACAT	CAAGGTGCCC	AAGAGCCGCT	ACCTGGGCTA	CATCAAGGAC	1200
TACGAGGGAG	CGACGCTGAT	GGAGTGTGAG	CTGAATCCCC	GCATCCCCCTA	CACGGAGCTG	1260
TCCCACATCA	TCAAGAAGCA	GAAAGAGATC	ATCAAGAAGC	TGATTGAGCG	CAAACAGGCC	1320
CAGATCCGCA	AGGTCTACCC	GGGGCTCAGC	TGCTTCAAGG	AGGGCGTGAG	GCAGATCCCT	1380
GTGGAGAGCG	TTCTTGCCAT	TCGAGAGACA	GGCTGGAAGC	CATTGGGGAA	GGAGAAGGGG	1440
AAGGAGCTGA	AGGACCCCGA	CCAGCTCTAC	ACAACCCTCA	AAAACCTGCT	GGCCCAAATC	1500
AAGTCTCACC	CCAGTGCCCTG	GCCCTTCATG	GAGCCTGTGA	AGAAGTCGGA	GGCCCTTGAC	1560
TACTACGAGG	TCATCCGCTT	CCCCATTGAC	CTGAAGACCA	TGACTGAGCG	GCTGCGAAGC	1620
CGCTACTACG	TGACCCGGAA	GCTCTTTGTG	GCCGACCTGC	AGCGGGTCAT	CGCCAACCTG	1680
CGCGAGTACA	ACCCCCCGGA	CAGCGAGTAC	TGCCGCTGTG	CCAGCGCCCT	GGAGAGTTTC	1740
TTTACTTTCA	AGCTCAAGGA	GGGAGGCCCT	ATTGACAAGT	AGGCCCATCT	TTGGGCCGCA	1800
GCCCTGACCT	GGAAATGTCT	CACCTCGGAT	TCTGATCTGA	TCCTTAGGGG	GTGCCCTGGC	1860
CCACCGGACC	CGACTCAGCT	TGAGACACTC	CAGCCAAGGG	TCCTCCGGAC	CCGATCCTGC	1920
AGCTCTTTCT	GGACCTTCAG	GCACCCCAAA	GCGTGACAGT	CTGTCCAGC	CTTCACTGTG	1980
TGTGAGAGGT	CTCTGGGTT	GGGGCCAGC	CCCTCTAGAG	TAGCTGGTGG	CCAGGGATGA	2040
ACCTTGCCCCA	GCCGTGGTGG	CCCCAGGCC	TGGTCCCCAA	GAGCCCGGAA	TTC	2093

## (2) INFORMATION FOR SEQ ID NO:12:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9046 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CCTTGTTTGT	GTGCTAGGCT	GGGGGGGAGA	GAGGGCGAGA	GAGAGCGGGC	GAGAGTGGGC	60
AAGCAGGACG	CCGGGCTGAG	TGCTAACTGC	GGGACGCAGA	GAGTGCGGAG	GGGAGTCGGG	120
TCGGAGAGAG	GCGGCAGGGG	CCAGAACAGT	GGCAGGGGGC	CCGGGGCGCA	CGGGCTGAGG	180
CGACCCCCAG	CCCCCTCCCG	TCCGCACACA	CCCCCACC GC	GGTCCAGCAG	CCGGGCCGGC	240
GTCGACGCTA	GGGGGGACCA	TTACATAACC	CGCGCCCCGG	CCGTCTTCTC	CCGCCGCCGC	300
GGCGCCCGAA	CTGAGCCCGG	GGCGGGCGCT	CCAGCACTGG	CCGCCGGCGT	GGGGCGTAGC	360
AGCGGGCCGA	TTATTATTTT	GCGGAAAGGA	AGGCGAAGGA	GGGGAGCGCC	GGCGCGAGGA	420
GGGGCCGCCT	GCGCCCGCCG	CCGGAGCGGG	GCCTCCTCGG	TGGGCTCCGC	GTCGGCGCGG	480
GCGTGCGGGC	GGCGCTGCTC	GGCCCGGCC	CCTCGGCCCT	CTGGTCCGGC	CAGCTCCGCT	540
CCCGGCGTCC	TTGCCGCGCC	TCCGCCGGCC	GCCGCGCGAT	GTGAGGCGGC	GGCGCCAGCC	600
TGGCTCTCGG	CTCGGGCGAG	TTCTCTGCGG	CCATTAGGGG	CCGGTGCGGC	GGCGGCGCGG	660
AGCGCGGCGG	CAGGAGGAGG	GTTCGGAGGG	TGGGGGCGCA	GGCCCGGGAG	GGGGCACC GG	720
GAGGAGGTGA	GTGTCTCTTG	TCGCCTCCTC	CTCTCCCCC	TTTTTCGCCCC	CGCCTCCTTG	780
TGGCGATGAG	AAGGAGGAGG	ACAGCGCCGA	GGAGGAAGAG	GTTGATGGCG	GCGGCGGAGC	840
TCCGAGAGAC	CTCGGCTGGG	CAGGGGCCGG	CCGTGGCGGG	CCGGGGACTG	CGCCTCTAGA	900
GCCGCGAGTT	CTCGGGAATT	CGCCGCAGCG	GACCGGCCTC	GGCGAATTTG	TGCTCTTGTG	960
CCCTCCTCCG	GGCTTGGGCC	AGGCCGGCCC	CTCGCACTTG	CCCTTACCTT	TTCTATCGAG	1020
TCCGCAATCC	TCTCCAGCCA	CTGCGACCCG	GCGAAGAGAA	AAAGGAACTT	CCCCACCCC	1080
CTCGGGTGCC	GTCGGAGCCC	CCCAGCCAC	CCCTGGGTGC	GGCGCGGGGA	CCCCGGGCCG	1140
AAGAAGAGAT	TTCTTGAGGA	TTCTGGTTTT	CCTCGCTTGT	ATCTCCGAAA	GAATTAATAA	1200
TGGCCGAGAA	TGTGGTGGAA	CCGGGGCCGC	CTTCAGCCAA	GCGGCCTAAA	CTCTCATCTC	1260
CGGCCCTCTC	GGCGTCCGCC	AGCGATGGCA	CAGATTTTGG	CTCTCTATTT	GACTTGGAGC	1320
ACGACTTACC	AGATGAATTA	ATCAACTCTA	CAGAATTGGG	ACTAACCAAT	GGTGGTGATA	1380

TTAATCAGCT	TCAGACAAGT	CTTGGCATGG	TACAAGATGC	AGCTTCTAAA	CATAAACAGC	1440
TGTCAGAATT	GCTGCGATCT	GGTAGTTCCC	CTAACCTCAA	TATGGGAGTT	GGTGGCCAG	1500
GTCAAGTCAT	GGCCAGCCAG	GCCCAACAGA	GCAGTCCTGG	ATTAGGTTTG	ATAAATAGCA	1560
TGGTCAAAAG	CCCAATGACA	CAGGCAGGCT	TGACTTCTCC	CAACATGGGG	ATGGGCACTA	1620
GTGGACCAA	TCAGGGTCCT	ACGCAGTCAA	CAGGTATGAT	GAACAGTCCA	GTAAATCAGC	1680
CTGCCATGGG	AATGAACACA	GGGACGAATG	CGGGCATGAA	TCCTGGAATG	TTGGCTGCAG	1740
GCAATGGACA	AGGGATAATG	CCTAATCAAG	TCATGAACGG	TTCAATTGGA	GCAGGCCGAG	1800
GGCGACAGGA	TATGCAGTAC	CCAAACCCAG	GCATGGGAAG	TGCTGGCAAC	TTACTGACTG	1860
AGCCTCTTCA	GCAGGGCTCT	CCCCAGATGG	GAGGACAAAC	AGGATTGAGA	GGCCCCCAGC	1920
CTCTTAAGAT	GGGAATGATG	AACAACCCCA	ATCCTTATGG	TTCACCATAT	ACTCAGAATC	1980
CTGGACAGCA	GATTGGAGCC	AGTGGCCTTG	GTCTCCAGAT	TCAGACAAAA	ACTGTACTAT	2040
CAATAACTT	ATCTCCATTT	GCTATGGACA	AAAAGGCAGT	TCCTGGTGGA	GGAAATGCCA	2100
ACATGGGTCA	ACAGCCAGCC	CCGCAGGTCC	AGCAGCCAGG	TCTGGTGACT	CCAGTTGCCC	2160
AAGGGATGGG	TTCTGGAGCA	CATACAGCTG	ATCCAGAGAA	GCGCAAGCTC	ATCCAGCAGC	2220
AGCTTGTTCT	CCTTTTGCAT	GCTCACAAGT	GCCAGCGCCG	GGAACAGGCC	AATGGGGAAG	2280
TGAGGCAGTG	CAACCTTCCC	CACTGTGCGA	CAATGAAGAA	TGTCTTAAAC	CACATGACAC	2340
ACTGCCAGTC	AGGCAAGTCT	TGCCAAGTGG	CACACTGTGC	ATCTTCTCGA	CAAATCATTT	2400
CACACTGGAA	GAATTGTACA	AGACATGATT	GTCTGTGTG	TCTCCCCCTC	AAAAATGCTG	2460
GTGATAAGAG	AAATCAACAG	CCAATTTTGA	CTGGAGCACC	CGTTGGACTT	GGAAATCCTA	2520
GCTCTCTAGG	GGTGGGTCAA	CAGTCTGCCC	CCAACCTAAG	CACTGTTAGT	CAGATTGATC	2580
CCAGCTCCAT	AGAAAGAGCC	TATGCAGCTC	TTGGACTACC	CTATCAAGTA	AATCAGATGC	2640
CGACACAACC	CCAGGTGCAA	GCAAAGAACC	AGCAGAATCA	GCAGCCTGGG	CAGTCTCCCC	2700
AAGGCATGCG	GCCCATGAGC	AACATGAGTG	CTAGTCCTAT	GGGAGTAAAT	GGAGGTGTAG	2760
GAGTTCAAAC	GCCGAGTCTT	CTTTCTGACT	CAATGTGCGA	TTCAGCCATA	AATTCTCAAA	2820
ACCCAATGAT	GAGTGAATAA	CCAGTGTGTC	CCTCCCTGGG	TCCTATGCCA	ACAGCAGCTC	2880
AACCATCCAC	TACTGGAATT	CGGAAACAGT	GGCACGAAGA	TATTACTCAG	GATCTTCGAA	2940
ATCATCTTGT	TCACAAACTC	GTCCAAGCCA	TATTTCTTAC	GCCGGATCCT	GCTGCTTTAA	3000
AAGACAGACG	GATGGAAAAC	CTAGTTGCAT	ATGCTCGGAA	AGTTGAAGGG	GACATGTATG	3060
AATCTGCAAA	CAATCGAGCG	GAATACTACC	ACCTTCTAGC	TGAGAAAATC	TATAAGATCC	3120
AGAAAGAACT	AGAAGAAAAA	CGAAGGACCA	GACTACAGAA	GCAGAACATG	CTACCAAATG	3180
CTGCAGGCAT	GGTTCCAGTT	TCCATGAATC	CAGGGCCTAA	CATGGGACAG	CCGCAACCAG	3240
GAATGACTTC	TAATGGCCCT	CTACCTGACC	CAAGTATGAT	CCGTGGCAGT	GTGCCAAACC	3300
AGATGATGCC	TCGAATAACT	CCACAATCTG	GTTTGAATCA	ATTTGGCCAG	ATGAGCATGG	3360
CCCAGCCCCC	TATTGTACCC	CGGCAAACCC	CTCCTCTTCA	GCACCATGGA	CAGTTGGCTC	3420
AACCTGGAGC	TCTCAACCCG	CCTATGGGCT	ATGGGCCTCG	TATGCAACAG	CCTTCCAACC	3480
AGGGCCAGTT	CCTTCCTCAG	ACTCAGTCTC	CATCACAGGG	AATGAATGTA	ACAAATTCCT	3540
CTTTGGCTCC	GCTCCAGCGT	CAAGCTCCAG	TGTCTCAAGC	ACAAATGTCT	AGTTCTTCCT	3600
GCCCGGTGAA	CTCTCCTATA	ATGCCTCCAG	GGTCTCAGGG	GAGCCACATT	CACTGTCCCC	3660
AGCTTCCTCA	ACCAGCTCTT	CATCAGAATT	CACCCTCGCC	TGTACCTAGT	CGTACCCCCA	3720
CCCCTCACCA	TACTCCCCCA	AGCATAGGGG	CTCAGCAGCC	ACCAGCAACA	ACAATTCCAG	3780
CCCCTGTTCC	TACACCACCA	GCCATGCCAG	CTGGGCCACA	GTCCCAGGCT	CTACATCCCC	3840
CTCCAAGGCA	GACACCTACA	CCACCAACAA	CACAACCTTC	CCAACAAGTG	CAGCCTTCAC	3900
TTCTGCTGTC	ACCTTCTGCT	GACCAGCCCC	AGCAGCAGCC	TCGCTCACAG	CAGAGCACAG	3960
CAGCGTCTGT	TCCTACCCCA	AACGCACCGC	TGCTTCCTCC	GCAGCCTGCA	ACTCCACTTT	4020
CCCAGCCAGC	TGTAAGCATT	GAAGGACAGG	TATCAAATCC	TCCATCTACT	AGTAGCACAG	4080
AAGTGAATTC	TCAGGCCATT	GCTGAGAAGC	AGCCTTCCCA	GGAAGTGAAG	ATGGAGGCCA	4140
AAATGGAAGT	GGATCAACCA	GAACCAGCAG	ATACGCAGCC	GGAGGATATT	TCAGAGTCTA	4200
AAGTGAAGA	CTGTAAAATG	GAATCTACCG	AAACAGAAGA	GAGAAGCACT	GAGTTAAAAA	4260
CTGAAATAAA	AGAGGAGGAA	GACCAGCCAA	GTACTTCAGC	TACCCAGTCA	TCTCCGGCTC	4320
CAGGACAGTC	AAAGAAAAAG	ATTTTCAAAC	CAGAAGAACT	ACGACAGGCA	CTGATGCCAA	4380
CATTGGAGGC	ACTTTACCGT	CAGGATCCAG	AATCCCTTCC	CTTTCGTCAA	CCTGTGGACC	4440
CTCAGCTTTT	AGGAATCCCT	GATTACTTTG	ATATTGTGAA	GAGCCCCATG	GATCTTTCTA	4500
CCATTAAGAG	GAAGTTAGAG	ACTGGACAGT	ATCAGGAGCC	CTGGCAGTAT	GTGATGATA	4560
TTTGGCTTAT	GTTCATAAAT	GCCTGGTTAT	ATAACCGGAA	AACATCACGG	GTATACAAAT	4620
ACTGCTCCAA	GCTCTCTGAG	GTCTTTGAAC	AAGAAATTGA	CCCAGTGATG	CAAAGCCTTG	4680
GATACTGTTG	TGGCAGAAAG	TTGGAGTTCT	CTCCACAGAC	ACTGTGTTGC	TACGGCAAAC	4740
AGTTGTGCAC	AATACCTCGT	GATGCCACTT	ATTACAGTTA	CCAGAACAGG	TATCATTTCT	4800
GTGAGAAGTG	TTTCAATGAG	ATCCAAGGGG	AGAGCGTTTC	TTTGGGGGAT	GACCCTTCCC	4860
AGCCTCAAAC	TACAATAAAT	AAAGAACAAT	TTTCCAAGAG	AAAAAATGAC	ACACTGGATC	4920
CTGAAGCTGT	TGTTGAATGT	ACAGAGTGC	GAAGAAAGAT	GCATCAGATC	TGTGTCTTTC	4980
ACCATGAGAT	CATCTGGCCT	GCTGGATTCT	TCTGTGATGG	CTGTTTAAAG	AAAAGTGCAC	5040
GAAGTAGGAA	AGAAAATAAG	TTTCTGCTA	AAAGGTTGCC	ATCTACCAGA	CTTGGCACCT	5100
TTCTAGAGAA	TCGTGTGAAT	GACTTCTGTA	GGCGACAGAA	TCACCCTGAG	TCAGGAGAGG	5160

TCACGTGTTAG	AGTAGTTCAT	GCTTCTGACA	AAACCGTGGA	AGTAAAACCA	GGCATGAAAG	5220
CAAGGTTTGT	GGACAGTGGA	GAGATGGCAG	AATCCTTTCC	ATACCGAACC	AAAGCCCTCT	5280
TTGCCTTTGA	AGAAATTGAT	GGTGTTGACC	TGTGCTTCTT	TGGCATGCAT	GTTCAAGAGT	5340
ATGGCTCTGA	CTGCCCTCCA	CCCAACCAGA	GGAGAGTATA	CATATCTTAC	CTCGATAGTG	5400
TTCATTTCTT	CCGTCTCTAA	TGCTTGAGGA	CTGCAGTCTA	TCATGAAATC	CTAATTGGAT	5460
ATTTAGAATA	TGTCAAGAAA	TTAGGTTACA	CAACAGGGCA	TATTTGGGCA	TGTCCACCAA	5520
GTGAGGGAGA	TGATTATATC	TTCCATTGCC	ATCCTCCTGA	CCAGAAGATA	CCCAAGCCCA	5580
AGCGACTGCA	GGAATGGTAC	AAAAAATGC	TTGACAAGGC	TGTATCAGAG	CGTATTGTCC	5640
ATGACTACAA	GGATATTTTT	AAACAAGCTA	CTGACAGATAG	ATTAACAAGT	GCAAAGGAAT	5700
TGCCTTATTT	CGAGGGTGAT	TTCTGGCCCA	ATGTTCTGGA	AGAAAGCATT	AAGGAAGTGG	5760
AACAGGAGGA	AGAAGAGAGA	AAACGAGAGG	AAAACACCAG	CAATGAAAGC	ACAGATGTGA	5820
CCAAGGGAGA	CAGCAAAAAT	GCTAAAAAGA	AGAATAATAA	GAAAACCAGC	AAAAATAAGA	5880
GCAGCCTGAG	TAGGGGCAAC	AAGAAGAAAC	CCGGGATGCC	CAATGTATCT	AACGACCTCT	5940
CACAGAAACT	ATATGCCACC	ATGGAGAAGC	ATAAAGAGGT	CTTCTTTGTG	ATCCGCCTCA	6000
TTGCTGGCCC	TGCTGCCAAC	TCCCTGCCTC	CCATTGTTGA	TCCTGATCCT	CTCATCCCTT	6060
GCGATCTGAT	GGATGGTCGG	GATGCGTTTC	TCACGCTGGC	AAGGGACAAG	CACCTGGAGT	6120
TCTCTTCACT	CCGAAGAGCC	CAGTGGTCCA	CCATGTGCAT	GCTGGTGGAG	CTGCACACGC	6180
AGAGCCAGGA	CCGCTTTGTC	TACACCTGCA	ATGAATGCAA	GCACCATGTG	GAGACACGCT	6240
GGCACTGTAC	TGTCTGTGAG	GATTATGACT	TGTGTATCAC	CTGCTATAAC	ACTAAAAACC	6300
ATGACCACAA	AATGGAGAAA	CTAGGCCTTG	GCTTAGATGA	TGAGAGCAAC	AACCAGCAGG	6360
CTGCAGCCAC	CCAGAGCCCA	GGCGATTCTC	GCCGCTGAG	TATCCAGCGC	TGCATCCAGT	6420
CTCTGGTCCA	TGCTTGCCAG	TGTCGGAAATG	CCAATTGCTC	ACTGCCATCC	TGCCAGAAGA	6480
TGAAGCGGGT	TGTGCAGCAT	ACCAAGGGTT	GCAAACGGAA	AACCAATGGC	GGGTGCCCCA	6540
TCTGCAAGCA	GCTCATTGCC	CTCTGCTGCT	ACCATGCCAA	GCACTGCCAG	GAGAACAAT	6600
GCCCGGTGCC	GTTCTGCCTA	AACATCAAGC	AGAAGCTCCG	GCAGCAACAG	CTGCAGCACC	6660
GACTACAGCA	GGCCCAAATG	CTTCGCAGGA	GGATGGCCAG	CATGCAGCGG	ACTGGTGTGG	6720
TTGGGCAGCA	ACAGGGCCCTC	CCTTCCCCCA	CTCCTGCCAC	TCCAACGACA	CCAACCTGGCC	6780
AACAGCCAAC	CACCCCGCAG	ACGCCCCAGC	CCACTTCTCA	GCCTCAGCCT	ACCCCTCCCA	6840
ATAGCATGCC	ACCCTACTTG	CCCAGGACTC	AAGCTGCTGG	CCCTGTGTCC	CAGGGTAAGG	6900
CAGCAGGCCA	GGTGACCCCT	CCAACCCCTC	CTCAGACTGC	TCAGCCACCC	CTTCCAGGGC	6960
CCCCACCTAC	AGCAGTGGAA	ATGGCAATGC	AGATTCAGAG	AGCAGCGGAG	ACGCAGCGCC	7020
AGATGGCCCA	CGTGCAAAAT	TTTCAAAGGC	CAATCCAACA	CCAGATGCCC	CCGATGCTC	7080
CCATGGGGCC	CATGGGTATG	AACCCACCTC	CCATGACCAG	AGGTCCCAGT	GGGCATTTGG	7140
AGCCAGGGAT	GGGACCGACA	GGGATGCAGC	AACAGCCACC	CTGGAGCCAA	GGAGGATTGC	7200
CTCAGCCCCA	GCAACTACAG	TCTGGGATGC	CAAGGCCAGC	CATGATGTCA	GTGGCCAGC	7260
ATGGTCAACC	TTTGAACATG	GCTCCACAAC	CAGGATTGGG	CCAGGTAGGT	ATCAGCCCAC	7320
TCAAACCAGG	CACTGTGTCT	CAACAAGCCT	TACAAAACCT	TTTGCGGACT	CTCAGGTCTC	7380
CCAGCTCTCC	CCTGCAGCAG	CAACAGGTGC	TTAGTATCCT	TCACGCCAAC	CCCCAGCTGT	7440
TGGCTGCATT	CATCAAGCAG	CGGGCTGCCA	AGTATGCCAA	CTCTAATCCA	CAACCCATCC	7500
CTGGGCAGCT	TGGCATGCCC	CAGGGGCAGC	CAGGGCTACA	GCCACCTACC	ATGCCAGGTC	7560
AGCAGGGGGT	CCACTCCAAT	CCAGCCATGC	AGAACATGAA	TCCAATGCAG	GCGGGCGTTC	7620
AGAGGGCTGG	CCTGCCCCAG	CAGCAACCAC	AGCAGCAACT	CCAGCCACCC	ATGGGAGGGA	7680
TGAGCCCCCA	GGCTCAGCAG	ATGAACATGA	ACCACAACAC	CATGCCTTCA	CAATTCCGAG	7740
ACATCTTGAG	ACGACAGCAA	ATGATGCAAC	AGCAGCAGCA	ACAGGGAGCA	GGGCCAGGAA	7800
TAGGCCCTGG	AATGGCCAAC	CATAACCACT	TCCAGCAACC	CCAAGGAGTT	GCTTACCCAC	7860
CACAGCCGCA	GCAGCGGATG	CAGCATCACA	TGCAACAGAT	GCAACAAGGA	AATATGGGAC	7920
AGATAGGCCA	GCTTCCCCAG	GCCTTGGGAG	CAGAGGCAGG	TGCCAGTCTA	CAGGCCTATC	7980
AGCAGCGACT	CCTTCAGCAA	CAGATGGGGT	CCCCTGTTCA	GCCCAACCCC	ATGAGCCCCC	8040
AGCAGCATAT	GCTCCCAAAT	CAGGCCCACT	CCCCACACCT	ACAAGGCCAG	CAGATCCCTA	8100
ATTCTCTCTC	CAATCAAGTG	CGCTCTCCCC	AGCCTGTCCC	TTCTCCACGG	CCACAGTCCC	8160
AGCCCCCCCC	CTCCAGTCTT	TCCCCAAGGA	TGCAGCCTCA	GCCTTCTCCA	CACCACGTTT	8220
CCCCACAGAC	AAGTTCCCCA	CATCCTGGAC	TGGTAGCTGC	CCAGGCCAAC	CCCATGGAAC	8280
AAGGGCATT	TGCCAGCCCC	GACCAGAATT	CAATGCTTTC	TCAGCTTGCT	AGCAATCCAG	8340
GCATGGCAAA	CCTCCATGGT	GCAAGCGCCA	CGGACCTGGG	ACTCAGCACC	GATAACTCAG	8400
ACTTGAATTC	AAACCTCTCA	CAGAGTACAC	TAGACATACA	CTAGAGACAC	CTTGTATTTT	8460
GGGAGCAAAA	AAATTATTTT	CTCTTAACAA	GACTTTTTGT	ACTGAAAACA	ATTTTTTTGA	8520
ATCTTTCTGA	GCCTAAAAGA	CAATTTTCTT	TGGAACACAT	AAGAACTGTG	CAGTAGCCGT	8580
TTGTGGTTTA	AAGCAAAACAT	GCAAGATGAA	CCTGAGGGAT	GATAGAATAC	AAAGAATATA	8640
TTTTTGTAT	GGGCTGGTTA	CCACCAGCCT	TTCTTCCCCT	TTGTGTGTGT	GGTTCAAGTG	8700
TGCACTGGGA	GGAGGCTGAG	GCCTGTGAAG	CCAAACAATA	TGCTCCTGCC	TTGCACCTCC	8760
AATAGGTTTT	ATTATTTTTT	TTAAATTAAT	GAACATATGT	AATATTAATG	AACATATGTA	8820
ATATTAATAG	TTATTATTTA	CTGGTGCAGA	TGGTTGACAT	TTTTCCCTAT	TTTCCCTACT	8880
TTATGGAAGA	GTTAAACAT	TTCTAAACCA	GAGGACAAAA	GGGGTTAATG	TTACTTTGAA	8940

ATTACATTCT	ATATATATAT	AAATATATAT	AAATATATAT	TAAAATACCA	GTTTTTTTTC	9000
TCTGGGTGCA	AAGATGTTCA	TTCTTTTAAA	AAATGTTTAA	AAAAAA		9046

## (2) INFORMATION FOR SEQ ID NO:13:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7326 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

ATGGCCGAGA	ACTTGCTGGA	CGGACCGCCC	AACCCCAAAC	GAGCCAAACT	CAGCTCGCCC	60
GGCTTCTCCG	CGAATGACAA	CACAGATTTT	GGATCATTGT	TTGACTTGGA	AAATGACCTT	120
CCTGATGAGC	TGATCCCCAA	TGGAGAATTA	AGCCTTTTAA	ACAGTGGGAA	CCTTGTTCCA	180
GATGCTGCGT	CCAAACATAA	ACAACGTGCA	GAGCTTCTTA	GAGGAGGCAG	CGGCTCTAGC	240
ATCAACCCAG	GGATAGGCAA	TGTGAGTGCC	AGCAGCCCTG	TGCAACAGGG	CCTTGGTGGC	300
CAGGCTCAGG	GGCAGCCGAA	CAGTACAAAC	ATGGCCAGCT	TAGGTGCCAT	GGGCAAGAGC	360
CCTCTGAACC	AAGGAGACTC	ATCAACACCC	AACCTGCCCA	AACAGGCAGC	CAGCACCTCT	420
GGGCCCCACTC	CCCCTGCCTC	CCAAGCACTG	AATCCACAAG	CACAAAAGCA	AGTAGGGCTG	480
GTGACCAAGTA	GTCCCTGCCAC	ATCACAGACT	GGACCTGGGA	TCTGCATGAA	TGCTAACTTC	540
AACCAGACCC	ACCCAGGCCT	TCTCAATAGT	AACCTCTGGCC	ATAGCTTAAT	GAATCAGGCT	600
CAACAAGGGC	AAGCTCAAGT	CATGAATGGA	TCTCTTGGGG	CTGCTGGAAG	AGGAAGGGGA	660
GCTGGAATGC	CCTACCCTGC	TCCAGCCATG	CAGGGGGCCA	CAAGCAGTGT	GCTGGCGGAG	720
ACCTTGACAC	AGGTTTCCCC	ACAAATGGCT	GGCCATGCTG	GACTAAATAC	AGCACAGGCA	780
GGAGGCATGA	CCAAGATGGG	AATGACTGGT	ACCACAAGTC	CATTTGGACA	ACCCTTTAGT	840
CAAACTGGAG	GGCAGCAGAT	GGGAGCCACT	GGAGTGAACC	CCCAGTTAGC	CAGCAAACAG	900
AGCATGGTCA	ATAGTTTACC	TGCTTTTCTT	ACAGATATCA	AGAATACTTC	AGTCACCACT	960
GTGCCAAATA	TGTCCCAGTT	GCAAACATCA	GTGGGAATTG	TACCCACACA	AGCAATTGCA	1020
ACAGGCCCCA	CAGCAGACCC	TGAAAAACGC	AAACTGATAC	AGCAGCAGCT	GGTCTACTG	1080
CTTCATGCCC	ACAAATGTCA	GAGACGAGAG	CAAGCAAATG	GAGAGGTTCT	NGCCTGTTCT	1140
CTCCCACACT	GTCGAACCAT	GAAAAACGTT	TTGAATCACA	TGACACATTG	TCAGGCTCCC	1200
AAAGCCTGCC	AAGTTGCCCA	TTGTGCATCT	TCACGACAAA	TCATCTCTCA	TTGGAAGAAC	1260
TGCACACGAC	ATGACTGTCC	TGTTTGCCCT	CCTTTGAAAA	ATGCCAGTGA	CAAGCGAAAC	1320
CAACAAACCA	TCCTGGGATC	TCCAGCTAGT	GGAAATCAAA	ACACAATTGG	TTCTGTTGGT	1380
GCAGGGCAAC	AGAATGCCAC	TTCTTAAAGT	AACCCAAATC	CCATAGACCC	CAGTTCCATG	1440
CAGCGGGCCT	ATGCTGCTCT	AGGACTCCCC	TACATGAACC	AGCCTCAGAC	GCAGCTGCAG	1500
CCTCAGGTTT	CTGGCCAGCA	ACCAGCACAG	CCTCCAGCCC	ACCAGCAGAT	GAGGACTCTC	1560
AATGCCCTAG	GAAACAACCC	CATGAGTGTC	CCAGCAGGAG	GAATAACAAC	AGATCAACAG	1620
CCACCAAACT	TGATTTTACA	ATCAGCTCTT	CCAACCTCCT	TGGGGGCTAC	CAATCCACTG	1680
ATGAATGATG	GTTCAAACCT	TGGTAACATT	GGAAGCCTCA	GCACGATACC	TACAGCAGCG	1740
CCTCCTTCCA	GCACTGGTGT	TCGAAAAGGC	TGGCATGAAC	ATGTGACTCA	GGACCTACGG	1800
AGTCATCTAG	TCCATAAACT	CGTTCAAGCC	ATCTTCCCAA	CTCCAGACCC	TGCAGCTCTG	1860
AAAGATCGCC	GCATGGAGAA	CCTGGTTGCC	TATGCTAAGA	AAGTGGAGGG	AGACATGTAT	1920
GAGTCTGCTA	ATAGCAGGGA	TGAATACTAT	CATTTATTAG	CAGAGAAAAT	CTATAAAATA	1980
CAAAAAGAAC	TAGAAGAAAA	GCGGAGGACA	CGTTTACATA	AGCAAGGCAT	CCTGGGTAAC	2040
CAGCCAGCTT	TACCAGCTTC	TGGGGCTCAG	CCCCCTGTGA	TTCCACCAGC	CCAGTCTGTA	2100
AGACCTCCAA	ATGGGCCCCCT	GCCTTTGCCA	GTGAATCGCA	TGCAGGTTTC	TCAAGGGATG	2160
AATTCATTTA	ACCCAATGTC	CCTGGGAAAC	GTCCAGTTGC	CACAGGCACC	CATGGGACCT	2220
CGTGCAGCCT	CCCCTATGAA	CCACTCTGTG	CAGATGAACA	GCATGGCCTC	AGTTCGGGGT	2280
ATGGCCATTT	CTCCTTCACG	GATGCCTCAG	CCTCCAAATA	TGATGGGCAC	TGATGCCAAC	2340
AACATTATGG	CCCAGGCACC	TACTCAGAAC	CAGTTTCTGC	CACAGAACCA	GTTTCCATCA	2400
TCCAGTGGGG	CAATGAGTGT	GAACAGTGTG	GGCATGGGGC	AACCAGCAGC	CCAGGCAGGT	2460
GTTTCACAGG	GTCAGGAACC	TGGAGCTGCT	CTCCCTAACC	CTCTGAACAT	GCTGGCACCC	2520
CAGGCCAGCC	AGCTGCCTTG	CCCACCAAGT	ACACAGTCAC	CATTGCACCC	GACTCCACCT	2580
CCTGCTTCCA	CAGCTGCTGG	CATGCCCTCT	CTCCAACATC	CAACGGCACC	AGGAATGACC	2640
CCTCCTCAGC	CAGCAGCTCC	CAGTCAGCTC	TCTCATCTGG	TGTCATCTGG	GCAGACTCCT	2700
ACCCCAACTC	CTGGCTCAGT	GCCCAGCGCT	GCCCAAAACAC	AGAGTACCCC	TACAGTCCAG	2760
GCAGCAGCAC	AGGCTCAGGT	GACTCCACAG	CCTCAGACCC	CAGTGCAGCC	ACCATCTGTG	2820
GCTACTCCTC	AGTCATCACA	GCAGCAACCA	ACGCCTGTGC	ATACTCAGCC	ACCTGGCACA	2880
CCGCTTTCTC	AGGCAGCAGC	CAGCATTGAT	AATAGAGTCC	CTACTCCCTC	CACTGTGACC	2940

AGTGCTGAAA	CCAGTTCCCA	GCAGCCAGGA	CCCGATGTGC	CCATGCTGGA	AATGAAGACA	3000
GAGGTGCAGA	CAGATGATGC	TGAGCCTGAA	CCTACTGAAT	CCAAGGGGGA	ACCTCGGTCT	3060
GAGATGATGG	AAGAGGATTT	ACAAGGTTCT	TCCCAAGTAA	AAGAAGAGAC	AGATACGACA	3120
GAGCAGAAGT	CAGAGCCAAT	GGAAGTAGAA	GAAAAGAAAC	CTGAAGTAAA	AGTGGAAGCT	3180
AAAGAGGAAG	AAGAGAACAG	TTCGAACGAC	ACAGCCTCAC	AATCAACATC	TCCTTCCAG	3240
CCACGCAAAA	AAATCTTTAA	ACCCGAGGAG	CTACGCCAGG	CACTTATGCC	AACTCTAGAA	3300
GCACTCTATC	GACAGGACCC	AGAGTCTTTG	CCTTTTCGTC	AGCCTGTAGA	TCCTCAGCTC	3360
CTAGGAATCC	CAGATTATTT	TGATATAGTG	AAGAATCCTA	TGGACCTTTC	TACCATCAAA	3420
CGAAAGCTGG	ACACAGGGCA	ATATCAAGAA	CCCTGGCAGT	ATGTGGATGA	TGTCAGGCTT	3480
ATGTTCAACA	ATGCGTGGCT	ATATAATCGT	AAAACGTCCC	GTGTATATAA	ATTTTGCAGT	3540
AAACTTGCAG	AGGTCTTTGA	ACAAGAAATT	GACCCTGTCA	TGCAGTCTCT	TGGATATTGC	3600
TGTGGACGAA	AGTATGAGTT	CTCCCCACAG	ACTTTGTGCT	GTTACGGAAA	GCAGCTGTGT	3660
ACAATTCCTC	GTGATGCAGC	CTACTACAGC	TATCAGAATA	GGTATCATTT	CTGTGGGGAAG	3720
TGTTTCACAG	AGATCCAGGG	CGAGAATGTG	ACCCTGGGTG	ACGACCCTTC	CCAACCTCAG	3780
ACGACAATTT	CCAAGGATCA	ATTTGAAAAG	AAGAAAAATG	ATACCTTAGA	TCCTGAACCT	3840
TTTGTGTGCT	GCAAAGAGTG	TGGCCGGAAG	ATGCATCAGA	TTTGTGTTCT	ACACTATGAC	3900
ATCATTGTGG	CTTCAGGTTT	TGTGTGTGAC	AACTGTTTGA	AGAAAACTGG	CAGACCTCGG	3960
AAAGAAAACA	AATTCAGTGC	TAAGAGGCTG	CAGACCACAC	GATTGGGAAA	CCACTTAGAA	4020
GACAGAGTGA	ATAAGTTTTT	GCGGCGCCAG	AATCACCCTG	AAGCTGGGGA	GGTTTTTGTG	4080
AGAGTGGTGG	CCAGCTCAGA	CAAGACTGTG	GAGGTCAAGC	CGGGAATGAA	GTCAAGGTTT	4140
GTGGATTCTG	GAGAGATGTC	GGAATCTTTC	CCATATCGTA	CCAAAGCACT	CTTTGCTTTT	4200
GAGGAGATCG	ATGGAGTCGA	TGTGTGCTTT	TTTGGGATGC	ATGTGCAAGA	TACGGCTCTG	4260
ATTGCCCCCC	ACCAAATACA	AGGCTGTGTA	TACATATCTT	ATCTGGACAG	TATTCATTTT	4320
TTCCGGCCCC	GCTGCCTCCG	GACAGCTGTT	TACCATGAGA	TCCTCATCGG	ATATCTCGAG	4380
TATGTGAAGA	AATTGGTGTG	TGTGACAGCA	CATATTTGGG	CCTGTCCCCC	AAGTGAAGGA	4440
GATGACTATA	TCTTTCATTG	CCACCCCCCT	GACCAGAAAA	TCCCCAAACC	AAAACGACTA	4500
CAGGAGTGGT	ACAAGAAGAT	GCTGGACAAG	GCCTTTGCAG	AGAGGATCAT	TAACGACTAT	4560
AAGGACATCT	TCAAACAAGC	GAACGAAGAC	AGGCTCAGCA	GTGCCAAGGA	GTTGCCCTAT	4620
TTTGAAGGAG	ATTTCTGGCC	TAATGTGTTG	GAAGAAAGCA	TTAAGGAACT	AGAACAAGAA	4680
GAAGAAGAAA	GGAAAAAAGA	AGAGAGTACT	GCAGCGAGTG	AGACTCCTGA	GGGCAGTCAG	4740
GGTGACAGCA	AAAATGCGAA	GAAAAAGAAC	AACAAGAAAG	CCAACAAAAA	CAAAAGCAGC	4800
ATTAGCCGCG	CCAACAAGAA	GAAGCCCAGC	ATGCCCAATG	TTTCCAACGA	CCTGTGCGAG	4860
AAGCTGTATG	CCACCATGGA	GAAGCACAAG	GAGGTATTCT	TTGTGATTCA	TCTGCATGCT	4920
GGGCCTGTTA	TCAGCACTCA	GCCCCCCTCA	GTGGACCCTG	ATCCTCTGCT	TAGCTGTGAC	4980
CTCATGGATG	GGCGAGATGC	CTTCCTCACC	CTGGCCAGAG	ACAAGCACTG	GGAAATCTCT	5040
TCCTTACGCC	GCTCCAAATG	GTCCACTCTG	TGCATGCTGG	TGGAGCTGCA	CACACAGGGC	5100
CAGGAACGCT	TTGTTTATAC	CTGCAATGAG	TGCAAAACACC	ATGTGGAAAC	ACGCTGGCAC	5160
TGCACTGTGT	GTGAGGACTA	TGACCTTTGT	ATCAATTGCT	ACAACACAAA	GAGCCACACC	5220
CATAAGATGG	TGAAGTGGGG	GCTAGGCCTA	GATGATGAGG	GCAGCAGTCA	GGGTGAGCCA	5280
CAGTCCAAGT	GCCCCCAGGA	ATCCCGGCGT	CTCAGCATCC	AGCGCTGCAT	CCAGTCCCTG	5340
GTGCATGCCT	GCCAGTGTG	CAATGCCAAC	TGCTCACTGC	CGTCTTGCCA	GAAGATGAAG	5400
CGAGTCGTGC	AGCACACCAA	GGGCTGCAAG	CGCAAGACTA	ATGGAGGATG	CCCAGTGTGC	5460
AAGCAGCTCA	TTGCTCTTTG	CTGCTACCCAC	GCCAAACACT	GCCAAGAAAA	TAAATGCCCT	5520
GTGCCCTTCT	GCCTCAACAT	CAAACATAAC	GTCCGCCAGC	AGCAGATCCA	GCAGTGCCTG	5580
CAGCAGGCTC	AGCTCATGCG	CCGGCGAATG	GCAACCATGA	ACACCCGCAA	TGTGCCTCAG	5640
CAGAGTTTGC	CTTCTCCTAC	CTCAGCACCA	CCCGGGACTC	CTACACAGCA	GCCCAGCACA	5700
CCCCAAACAC	CACAGCCCCC	AGCCCAGCCT	CAGCCTTCAC	CTGTTAACAT	GTCACCAGCA	5760
GGCTTCCCTA	ATGTAGCCCC	GACTCAGCCC	CCAACAATAG	TGTCTGCTGG	GAAGCCTACC	5820
AACCAGGTGC	CAGCTCCCCC	ACCCCCTGCC	CAGCCCCCAC	CTGCAGCAGT	AGAAGCAGCC	5880
CGGCAAATTG	AACGTGAGGC	CCAGCAGCAG	CAGCACCTAT	ACCGAGCAAA	CATCAACAAT	5940
GGCATGCCCC	CAGGACGTGA	CGGTATGGGG	ACCCCAGGAA	GCCAAATGAC	TCCTGTGGGG	6000
CTGAATGTGC	CCCGTCCCAA	CCAAGTCAGT	GGGCCTGTCA	TGTCTAGTAT	GCCACCTGGG	6060
CAGTGGCAGC	AGGCACCCAT	CCCTCAGCAG	CAGCCGATGC	CAGGCATGCC	CAGGCCTGTA	6120
ATGTCCATGC	AGGCCCAGGC	AGCAGTGGCT	GGGCCACGGA	TGCCCAATGT	GCAGCCAAAC	6180
AGGAGCATCT	CGCCAAGTGC	CCTGCAAGAC	CTGCTACGGA	CCCTAAAGTC	ACCCAGCTCT	6240
CCTCAGCAGC	AGCAGCAGGT	GCTGAACATC	CTTAAATCAA	ACCCACAGCT	AATGGCAGCT	6300
TTCATCAAAC	AGCGCACAGC	CAAGTATGTG	GCCAATCAGC	CTGGCATGCA	GCCCCAGCCC	6360
GGACTTCAAT	CCCAGCCTGG	TATGCAGCCC	CAGCCTGGCA	TGCACCAGCA	GCCTAGTTTG	6420
CAAAACCTGA	ACGCAATGCA	AGCTGGTGTG	CCACGGCCTG	GTGTGCCTCC	ACCACAACCA	6480
GCAATGGGAG	GCCTGAATCC	CCAGGGACAA	GCTCTGAACA	TCATGAACCC	AGGACACAAC	6540
CCCAACATGA	CAAACATGAA	TCCACAGTAC	CGAGAAATGG	TGAGGAGACA	GCTGCTACAG	6600
CACCAGCAGC	AGCAGCAGCA	ACAGCAGCAG	CAGCAGCAGC	AACAACAAAA	TAGTGCCAGC	6660
TTGGCCGGGG	GCATGGCGGG	ACACAGCCAG	TTCCAGCAGC	CACAAGGACC	TGGAGGTTAT	6720

GCCCCAGCCA	TGCAGCAGCA	ACGCATGCAA	CAGCACCTCC	CCATCCAGGG	CAGCTCCATG	6780
GGCCAGATGG	CTGCTCCAAT	GGGACAACCT	GGCCAGATGG	GGCAGCCTGG	GCTAGGGGCA	6840
GACAGCACCC	CTAATATCCA	GCAGGCCCTG	CAGCAACGGA	TTCTGCAGCA	GCAGCAGATG	6900
AAGCAACAAA	TTGGGTCAAC	AGGCCAGCCG	AACCCCATGA	GCCCCAGCA	GCACATGCTC	6960
TCAGGACAGC	CACAGGCCTC	ACATCTCCCT	GGCCAGCAGA	TCGCCACATC	CCTTAGTAAC	7020
CAGGTGCGAT	CTCCAGCCCC	TGTGCAGTCT	CCACGGCCCC	AATCCCAACC	TCCACATTCC	7080
AGCCCGTCAC	CACGGATACA	ACCCAGCCT	TCACCACACC	ATGTTTCACC	CCAGACTGGA	7140
ACCCCTCACC	CTGGACTCGC	AGTCACCATG	GCCAGCTCCA	TGGATCAGGG	ACACCTGGGG	7200
AACCCTGAAC	AGAGTGCAAT	GCTCCCCCAG	CTGAATACCC	CCAACAGGAG	CGCACTGTCC	7260
AGTGAAGTGT	CCCTGGTTGG	TGATACCACG	GGAGACACAC	TAGAAAAGTT	TGTGGAGGGT	7320
TTGTAG						7326

## (2) INFORMATION FOR SEQ ID NO:14:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2499 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

TCACTTGTCA	ATTAATCCAG	CTTCCTTAAT	TTTACTGAAG	AAGAATTTCT	CCAGGATATT	60
GGCACATTTG	TAGTATTCAC	TCTCAGGGGC	GTTGTAATCT	TTGCAATTGG	TAAAGACTCG	120
CTGTAAGTCT	GCCATGAATA	ATTTCTTAGA	CACGTAGTAC	CTATTCTTGA	GGCGTTCACT	180
CATGGTTTTG	AGATCCATGG	GGGACCTTAT	AACTTCATAA	TATCCTGGAG	CTTCTGTTCT	240
CTTCACAGGT	TCCATGAAGG	GCCAAGCGCT	TTGATGGCTC	TTACCTGCT	GGAGGATGCT	300
CTTGAGCGTG	CTGTAAAGCT	GGTCAGGGTC	TCTGGGCTCT	TTACTTTTCT	CTTTTCCACT	360
CGGTTTCCAG	CCTGTCTCTC	TAATTCCAGG	AATGCTTTCT	ATAGGAATCT	GTCGAATCC	420
ATCTTTAAAA	CATGAAAGTC	CAGGGTAAAC	TTTTCGAATT	TGTGCCTGTT	TTCTTTCAAT	480
CAGTTTTTTA	ATTATCTCCT	TCTGCTTTTT	AATGATGACA	GAAAATTCTG	TGTACGGGAT	540
CCGTGGATTT	AGCTCACATC	CCATTAAAGT	GGCTCCTTCA	TAATCCTTGA	TATAGCCAAC	600
ATATTTGGTT	TTAGGTATTT	TAATTTCTTT	GGAGAAACCC	TGTTTCTTAA	AGTATCCAAT	660
TGCATATTCA	TCTGCATATG	TGAGGAAGTT	CAGGATGTCA	TGCTTTATGT	GATATTCTTT	720
CAAATGATTC	ATCAGGTGTG	TTCCATAGCC	CTTGACTTGC	TCATTTGAGG	TTACAGCACA	780
GAAGACAATC	TCTGTGAATC	CTTGAGATGG	GAACATACGG	AAACAGATAC	CACCAATAAC	840
ACGGCCATCT	TTAATTAAAG	CAAGGGTTTT	GTGTTTCGGG	TCAAAGACGA	GCCGTGTGAT	900
GTATTCTTTT	GGCATTCCGG	GCAGCTGGTG	GGAGAAAACG	TTCTGTAGGC	CAACCAGCCA	960
CATCAGGATC	TTCTTGTTTG	GTTTCTGGTT	GAGGGAATTG	CCAACCACGT	GAAATTCAAT	1020
TACACCCCTG	CGCTCTTCCA	ACCTTGCCGC	CTCATCCCTG	GCCGAGTGTG	CTGACAGAAA	1080
ATTGGTCTCT	GGTCCAAGCA	TTGCTGCAGG	GTCCGTGATG	GTAGACATAA	CCTCGTTGAT	1140
TAATTCATC	CGAATATCCC	CCATAACTCG	GGGTTTCTTG	GCCTCCTCCA	GAACATGAGA	1200
ATCAGTCATT	TTCTCTTTT	CTCCTGGGTT	TGCTTCAAGT	CCAGAAGAGG	CTTTCAGGGC	1260
AGGACTGCTG	CTCCCTGCGT	TTGGCTGCTC	AAGGGAAGAT	GAGGTGGAAT	TGTATGAAAT	1320
TGTCCAGCC	ACAGGAGGTG	GATTGATAAC	TGTTTGGATG	CCTAGCTGGC	TGGTCTTGGA	1380
AGAGGCTGAG	AGAAAATCCT	GATCCAGAT	GGGAGAGTTT	TGACTATATA	CTTCTTCTTC	1440
TAGCATGGAC	AGAAATTTTG	GGAAATGAGT	GAGGATTAGA	GTTTCGTTTTT	CAAGAGGCAG	1500
TTTATCTTTT	TCCTGTCTTG	CTTGTTCCAG	GAGTTGTCGC	CTCATAACAG	TGAAGACCGA	1560
GCGAAGCAAT	GTTCTCCCAA	ACACCTGTGT	GGTTTCGTAC	CGAGGTAGAC	TGTCGCAGAA	1620
CTGTGGCAGG	TTGCAGTAAC	ACAGCCACCT	TGTGTAAGTC	TCTTTGTATC	CAGAAATATC	1680
ATCATTGGGA	GATCGCAGTC	TTCGTTGAGA	TGGTGCCTCC	AGATGCCAAT	AGTTGATGCG	1740
GTTTAGGAAC	ATTTTGGCA	ACTCAACTAT	TGTTTGCTT	TCTTTTGCTG	GCAGGTGACT	1800
AAATTTGTAC	TGCACAAAGT	TATTCACACC	CTGTTCAATG	CTAGGTTTTT	CAAATGGGGG	1860
TTTCTTTTCC	AAAGAGCCTT	CAACCACAGG	TTTTCTCTCT	TGTAAAATAG	ACTTTCTCAA	1920
GAGCTTAAAT	AGATAGAAAT	AAACTTGTTT	GGTATCTGCA	TCTTCTTCCT	TGTGGACACA	1980
GGTAAAGAGA	TATTCCACAT	CCAATACTAT	TCCCAGGAGT	CTGTTCAATT	CTTCTCTGTA	2040
CACATTCTCC	AGGTGGGAAA	CATGAGCAGC	TAGGGCATGG	CTACAACCTC	GACAGGATTG	2100
TGTTAGCTAG	ACAATTATTT	GCTGCAGGTC	GGCTCTGGGG	GGAGTGGGTG	AGGGGTATAG	2160
GTTTTTCCAG	CCATTACATT	TACAAGACTC	CTCGGCCCTG	CAGGCGGAGT	ACACTCCGAG	2220
TTTCTCCAGT	TTCTTGCCCC	GCGGAGCGGA	GCGTAGTTGC	GCTTTCTTCA	CGGCGATTCTG	2280
GGCCGAGCCA	CCGCCTCCCG	GTCCTTCGGC	CGTGCCCGCT	GCAGCCACTG	CCGTCGCCGG	2340
ACCGCAGGCG	CCCGAGCCCC	GCGCGGCAGC	GGCGCAGGGG	GAGCCCTGCG	GGGGCGCGGG	2400

CGGAAGCGCC GCAGGCTGCG GGGGCAGCGC CCCGGGCCCCG GCCCCTGCCC CGGCTCCTGC 2460  
 CCCGCAGCCG CCCGGCCCCG CCCCGCCAGC CTCGGACAT 2499

## (2) INFORMATION FOR SEQ ID NO:15:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2442 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

TCACCTTGTC	ATCAACCCTG	CTTCCTTAAT	TTTACTGAAG	AAGAACTTCT	CCAGGATGCT	60
GGCGCATTTG	TAGTACTCGC	TCTCGGGAGG	GTGTACTTCC	TTGCAGTTGG	TGAACACTCG	120
TTGCAAGTCC	GCCATGAATA	ACTTCTTAGA	CACATAGTAC	CTGTTCTTGA	GGCGTTCCT	180
CATGGTTTTT	AGATCCATGG	GGAACCTTAT	AACTTCATAA	TATCCCGGAG	CTTCTGTTCT	240
CTTCACTGGT	TCCATGAAAG	GCCAAGCATT	TGGATGGTTC	TTCACCTGCT	GCAGGATGTT	300
CTTGAGGGTG	CTGTAAACGT	GCTCAGGGTC	TTTGGGCTCT	TTACTTTTCT	CTTTTCCACT	360
TGGTTTCCAG	CCTGTCTCTC	TGATTCCAGG	AATGCTTTCT	ATAGGAATCT	GCCGAACCTC	420
ATCTTTGAAA	CACGAAAGTC	CAGGGTAGAC	TTTTCGAATC	TGGGCTTGTT	TTCTTTCTAT	480
CAGCTTTTAA	ATGATCTCCT	TCTGCTTTTT	AATGATGACA	GAGAACTCTG	TGTATGGGAT	540
CTGAGGGTTC	AGCTCACATC	CCATCAAAGT	GGCCCCTTCA	TAATCCTTGA	TGTAGCCAAC	600
ATATTTGGTT	TTAGGTATTT	TGATTTCTTT	GGAGAAACCC	TGCTTCTTGA	AATAGCCGAT	660
GGCATACTCA	TCTGCATATG	TGAGGAAGTT	GAGGATCTCG	TGCTTTATGT	GGTATTCTTT	720
GAGATGGTTC	ATCAGGTGGG	TTCCATAGCC	CTTGACTTGT	TCATTTGAGG	TTACTGCACA	780
GAAAACAATC	TCTGTGAATC	CCTGGGATGG	AAACATCCGG	AAACAGATAC	CACCAATGAC	840
ACGGCCATCT	TTAATTAAAG	CAAGGGTTTT	GTGTTTCGGG	TCAAAGACGA	GCCGTGTGAT	900
GTACTCTTTG	GGCATTCTGG	GCAGCTGGTG	GGAAAACACA	TTCTGGAGGC	CCACGAGCCA	960
CATCAGGATC	TTCTTGTTTG	GTTTCTGGTT	CAGGGAGTTG	CCCACCACGT	GGAATTC AAT	1020
GACACCCCTG	CGTTCTTCCA	GCCGTGCCGC	CTCATCTCTG	GCCGAATGGG	CTGACAGAAA	1080
ATTGGTCTCT	GGTCCAAGCA	TCCCTGCAGG	GTCTGTGATG	GTAGACATGA	CCTCATTGAT	1140
CAATTCACAG	GGAATATCCC	CCATCACTCG	AGATCTCTTG	GCCTCCTCGG	GAGCATGAGA	1200
GTGTGTTTCT	TTCTCTTTTT	CTCCCGGGTT	TGCTTCAAGC	CCAGAAGAGC	CTCTGCATCC	1260
AGGACTTGTT	CTCCCTCCAT	TGATCTGCTC	ATGGGAAGTT	GAATTTGAAC	TGAACAATGC	1320
TGTCCCAGTA	ACAGGAGGAC	TGATTACTGT	TTGGATTCTT	AGCGGGCTGG	TTCTGGAAGA	1380
GGCTGAGAGA	AAATCCTGAT	CCCAGATAGG	AGAATTTTGA	CTATACACTT	CTTCTTCCAA	1440
CATGGACAGA	AACTTTGGGA	AATGTGTGAG	GATAAGCGTG	CGTTTCTCAA	GAGGCAGTTT	1500
GTCTTTTTTC	TGTCTGGCTT	GTTCCAAGAG	CTGTCTCTTC	ATGATGGTGA	AGACCGAGCG	1560
AAGCAATGTT	CTCCCAAACA	CCTTTGTGGT	TTCGTACCGA	GGTAAGCTGT	CACAGAAGTG	1620
CGGTACATTG	CAGTAGCACA	ACCACCTTGT	GATGTTTTCC	TTGTATCCAG	AGATGTCATC	1680
ATTGGGAGAC	CGTAGTCTCC	GCTGAGATGG	AGCCTCCAGA	TGCCAGTAGT	TGATGCGGTT	1740
CAGAAACATC	TTGGCCAGCT	CGATCGTTGT	CTGCCTCTCT	TTCGATGGCA	AGTGACTAAA	1800
CTTGTAAGTG	ACGAAGTTGT	TCACACCTTG	TTCAATACTG	GGCTTCTCAA	ATGGCGGCTT	1860
CTTCTCCAAG	GAGCCTTCAA	CCACAGGTTT	TCCTCTTTGT	AAAATTGACT	TTCTCAAGAG	1920
CTTGAATAGG	TAGAAGTACA	CTTGTTTGGT	ATCTGCATCT	TCTTCTTTGT	GGACGCAGGT	1980
GAAGAGGTAC	TCCACATCCA	ACACAATTCC	CAGGAGTCTG	TCCATCTCTT	CCTCTGACAC	2040
ATTCTCCAAG	TGAGAAACGT	GAGCAGCAAG	GGCATGGCTA	CAGCTTCGAC	AGGATTCTGT	2100
CAAACCTGAC	ATTATCTGCT	GGAGGTCTCC	TCTTGGTGGA	GTAGGAGAGG	GGTTAGGGTT	2160
CTTCCAGCCA	TTGCATTTAC	AGGACTCCTC	TGCCTTGACG	GCGGAGTACA	CGCCGAGTTT	2220
CTCCAGCTTC	TTGCCCCGCG	GAGCAGAGCG	CAACTGCGCC	TTCTTCAACG	CGATCCGGGC	2280
CGAGCCGCGT	CCTCCCGGTC	CCTCGGCGGT	GCCCGCCGCG	GCCACCGGCG	TCGCTGGCCC	2340
GCAGGAAGCA	GAGCTCCCGG	CAGCGGTGGC	CAGGGTCCGG	GGGGAAACGT	GCGGGGGCGC	2400
GGGAGGCAGT	GCTGGGGACC	CGGCCCCGCC	AGCCTCGGCC	AT		2442

## (2) INFORMATION FOR SEQ ID NO:16:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CCCGCCAGCC TCGGACATGC

20

## (2) INFORMATION FOR SEQ ID NO:17:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

CCCGCCAGCC TCGGCCATGC

20

## (2) INFORMATION FOR SEQ ID NO:18:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2442 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

ATGGCCGAGG	CTGGCGGGGC	CGGGTCCCCA	GCACTGCCTC	CCGCGCCCCC	GCACGGTTCC	60
CCCCGGACCC	TGGCCACCGC	TGCCGGGAGC	TCTGCTTCCT	GCGGGCCAGC	GACGCCGGTG	120
GCCGCGGCGG	GCACCGCCGA	GGGACCGGGA	GGAGGCGGCT	CGGCCCGGAT	CGCCGTGAAG	180
AAGGCGCAGT	TGCGCTCTGC	TCCGCGGGCG	AAGAAGCTGG	AGAAACTCGG	CGTGTACTCC	240
GCCTGCAAGG	CAGAGGAGTC	CTGTAAATGC	AATGGCTGGA	AGAACCCTAA	CCCCTCTCCT	300
ACTCCACCAA	GAGGAGACCT	CCAGCAGATA	ATTGTCAGTT	TGACAGAATC	CTGTCGAAGC	360
TGTAGCCATG	CCCTTGCTGC	TCACGTTTCT	CACCTGGAGA	ATGTGTCAGA	GGAAGAGATG	420
GACAGACTCC	TGGGAATTGT	GTTGGATGTG	GAGTACCTCT	TCACCTGCGT	CCACAAAGAA	480
GAAGATGCAG	ATACCAAACA	AGTGTACTTC	TACCTATTCA	AGCTCTTGAG	AAAGTCAATT	540
TTACAAAGAG	GAAAACCTGT	GTTTGAAGGC	TCCTTGGAGA	AGAAGCCGCC	ATTTGAGAAG	600
CCCAGTATTG	AACAGGGTGT	GAACAACTTC	GTGCAGTACA	AGTTTAGTCA	CTTGCCATCG	660
AAAGAGAGGC	AGACAACGAT	CGAGCTGGCC	AAGATGTTTC	TGAACCGCAT	CAACTACTGG	720
CATCTGGAGG	CTCCATCTCA	GCGGAGACTA	CGGTCTCCCA	ATGATGACAT	CTCTGGATAC	780
AAGGAAAAC	ACACAAGGTG	GTTGTGCTAC	TGCAATGTAC	CGCAGTTCTG	TGACAGCTTA	840
CCTCGGTACG	AAACCACAAA	GGTGTTTGGG	AGAACATTGC	TTGCTCGGT	CTTCACCATC	900
ATGAGACGAC	AGCTCTTGGA	ACAAGCCAGA	CAGAAAAAAG	ACAAACTGCC	TCTTGAGAAA	960
CGCAGCCTTA	TCCTCACACA	TTTCCCAAAG	TTTCTGTCCA	TGTTGGAAGA	AGAAGTGTAT	1020
AGTCAAAATT	CTCCTATCTG	GGATCAGGAT	TTTCTCTCAG	CCTCTTCCAG	AACCAGCCCG	1080
CTAGGAATCC	AAACAGTAAT	CAGTCCTCCT	GTTACTGGGA	CAGCATTGTT	CAGTTCAAAT	1140
TCAACTTCCC	ATGAGCAGAT	CAATGGAGGG	AGAACAAGTC	CTGGATGCAG	AGGCTCTTCT	1200
GGGCTTGAAG	CAAACCCGGG	AGAAAAGAGG	AAAATGAACA	ACTCTCATGC	TCCCGAGGAG	1260
GCCAAGAGAT	CTCGAGTGAT	GGGGGATATT	CCCGTGGAA	TGATCAATGA	GGTCATGTCT	1320
ACCATCACAG	ACCCTGCAGG	GATGCTTGGA	CCAGAGACCA	ATTTTCTGTC	AGCCCATTCT	1380
GCCAGAGATG	AGGCGGCACG	GCTGGAAGAA	CGCAGGGGTG	TCATTGAATT	CCACGTGGTG	1440
GGCAACTCCC	TGAACCAGAA	ACCAAACAAG	AAGATCCTGA	TGTGGCTCGT	GGGCCTCCAG	1500
AATGTGTTTT	CCCACCAGCT	GCCCAGAATG	CCCAAAGAGT	ACATCACACG	GCTCGTCTTT	1560
GACCCGAAAC	ACAAAACCCT	TGCTTTAATT	AAAGATGGCC	GTGTCATTGG	TGGTATCTGT	1620
TTCCGGATGT	TTCCATCCCA	GGGATTCA	GAGATTGTTT	TCTGTGCACT	AACCTCAAAT	1680
GAACAAGTCA	AGGGCTATGG	AACCCACCTG	ATGAACCATC	TCAAAGAATA	CCACATAAAG	1740
CACGAGATCC	TCAACTTCCT	CACATATGCA	GATGAGTATG	CCATCGGCTA	TTTCAAGAAG	1800
CAGGGTTTCT	CCAAAGAAAT	CAAAATACCT	AAAACCAAAT	ATGTTGGCTA	CATCAAGGAT	1860
TATGAAGGGG	CCACTTTGAT	GGGATGTGAG	CTGAACCCCTC	AGATCCCATC	CACAGAGTTC	1920
TCTGTCAATC	TTAAAAAGCA	GAAGGAGATC	ATTAATAAAGC	TGATAGAAAG	AAAACAAGCC	1980
CAGATTTCGAA	AAGTCTACCC	TGGACTTTCTG	TGTTTCAAAG	ATGGAGTTCTG	GCAGATTCTCT	2040

ATAGAAAGCA	TTCCTGGAAT	CAGAGAGACA	GGCTGGAAAC	CAAGTGGAAA	AGAGAAAAGT	2100
AAAGAGCCCA	AAGACCCTGA	GCACGTTTAC	AGCACCTTCA	AGAACATCCT	GCAGCAGGTG	2160
AAGAACCATC	CAAATGCTTG	GCCTTTCATG	GAACCAGTGA	AGAGAACAGA	AGCTCCGGGA	2220
TATTATGAAG	TTATAAGGTT	CCCCATGGAT	CTGAAAACCA	TGAGTGAACG	CCTCAGGAAC	2280
AGGTACTATG	TGTCTAAGAA	GTTATTTCATG	GCGGACTTGC	AACGAGTGTT	CACCAACTGC	2340
AAGGAGTACA	ACCCTCCCGA	GAGCGAGTAC	TACAAATGCG	CCAGCATCCT	GGAGAAGTTC	2400
TTCTTCAGTA	AAATTAAGGA	AGCAGGGTTG	ATTGACAAGT	GA		2442

**What is claimed is:**

1. A purified protein designated P/CAF having a molecular weight of about 93,000 daltons as determined by sodium dodecyl sulfate polyacrylamide gel electrophoresis under reducing conditions and which acetylates histones.
2. The protein of claim 1 consisting of the amino acid sequence of SEQ ID NO:1.
3. The protein of claim 1 comprising the amino acid sequence of SEQ ID NO:2.
4. The protein of claim 1, which also binds to the amino acid sequence of SEQ ID NO:3 on a p300 cellular protein and to amino acid residues 1805-1854 of a CBP cellular protein (SEQ ID NO:9).
5. A fragment of the protein of claim 1 having histone acetyltransferase activity.
6. A polypeptide consisting of the amino acid sequence of SEQ ID NO: 2.
7. A fragment of the protein of claim 1 which binds to the amino acid sequence of SEQ ID NO: 3 on the p300 cellular protein and the amino acid sequence of SEQ ID NO:9 on the CBP cellular protein.
8. A polypeptide consisting of the amino acid sequence of SEQ ID NO:4.
9. A nucleic acid consisting of the nucleotide sequence of SEQ ID NO:10.
10. A nucleic acid having a nucleotide sequence which encodes the protein of claim 1.

11. A nucleic acid having a nucleotide sequence which encodes the protein of claim 2.
12. A nucleic acid having a nucleotide sequence which encodes the protein of claim 3.
13. A nucleic acid consisting of the nucleotide sequence which encodes the protein of claim 4.
14. A nucleic acid complementary to and which selectively hybridizes with the nucleic acid of claim 11 under stringent hybridization conditions.
15. A fragment of the nucleic acid of claim 9, which encodes a polypeptide that acetylates histones.
16. A fragment of the nucleic acid of claim 9, which encodes a polypeptide which binds to the amino acid sequence of SEQ ID NO:3 on the p300 cellular protein and the amino acid sequence of SEQ ID NO:9 on the CBP cellular protein.
17. A purified antibody which specifically binds the protein of claim 1.
18. A purified antibody which specifically binds the protein of claim 2.
19. A purified antibody which specifically binds the protein of claim 3.
20. A purified antibody which specifically binds the protein of claim 4.
21. An assay for screening substances for the ability to inhibit or stimulate the histone acetyltransferase activity of P/CAF comprising:
  - a) contacting the substance with a system in which histone acetylation by P/CAF can be determined;

b) determining the amount of histone acetylation by P/CAF in the presence of the substance; and

c) comparing the amount of histone acetylation by P/CAF in the presence of the substance with the amount of histone acetylation by P/CAF in the absence of the substance, a decreased or increased amount of histone acetylation by P/CAF in the presence of the substance indicating a substance that can inhibit or stimulate, respectively, the histone acetyltransferase activity of P/CAF.

22. An assay for screening substances for the ability to inhibit binding of P/CAF to p300/CBP comprising:

a) contacting the substance with a system in which the P/CAF binding of p300/CBP can be determined;

b) determining the amount of P/CAF binding of p300/CBP in the presence of the substance; and

c) comparing the amount of binding of P/CAF to p300/CBP in the presence of the substance with the amount of binding of P/CAF to p300/CBP in the absence of the substance, a decreased amount of binding of P/CAF to p300/CBP in the presence of the substance indicating a substance that can inhibit the ability to inhibit binding of P/CAF to p300/CBP.

23. The method of claim 22, wherein the system consists of a cell free reaction mixture comprising a fragment of the p300 protein comprising amino acid residues 1767-1816 (SEQ ID NO:3) and the protein of claim 4.

24. The method of claim 22, wherein the system consists of a cell free reaction mixture comprising a fragment of the CBP protein comprising amino acid residues 1805-1854 (SEQ ID NO:9) and the protein of claim 4.

25. The method of claim 22, wherein the system consists of a cell extract produced from cells producing both p300 and P/CAF.

26. An assay for screening substances for the ability to inhibit or stimulate the histone acetyltransferase activity of p300/CBP comprising:

- a) contacting the substance with a system in which histone acetylation by p300/CBP can be determined;
- b) determining the amount of histone acetylation by p300/CBP in the presence of the substance; and
- c) comparing the amount of histone acetylation by p300/CBP in the presence of the substance with the amount of histone acetylation by p300/CBP in the absence of the substance, a decreased or increased amount of histone acetylation by p300/CBP in the presence of the substance indicating a substance that can inhibit or stimulate, respectively, the histone acetyltransferase activity of p300/CBP.

27. An assay for screening substances for the ability to inhibit binding of a DNA-binding transcription factor to p300/CBP comprising:

- a) contacting the substance with a system in which the DNA-binding transcription factor binding of p300/CBP can be determined;
- b) determining the amount of DNA-binding transcription factor binding of p300/CBP in the presence of the substance; and
- c) comparing the amount of binding of DNA-binding transcription factor to p300/CBP in the presence of the substance with the amount of binding of DNA-binding transcription factor to p300/CBP in the absence of the substance, a decreased amount of binding of DNA-binding transcription factor to p300/CBP in the presence of the substance indicating a substance that can inhibit the ability to inhibit binding of DNA-binding transcription factor to p300/CBP.

28. The method of claim 27, wherein the system consists of a cell free reaction mixture comprising a DNA-binding transcription factor and p300/CBP.

29. The method of claim 27, wherein the system consists of a cell free reaction mixture comprising a fragment of the CBP protein comprising a DNA-binding transcription factor and p300/CBP.

30. The method of claim 27, wherein the system consists of a cell extract produced from cells producing both a DNA-binding transcription factor and p300/CBP.
31. The method of claim 27, wherein the DNA-binding transcription factor is selected from the group consisting of a nuclear hormone receptor, CREB, c-Jun/v-Jun, c-Myb/v-Myb, YY1, Sap-1a, c-Fos, MyoD and SRC-1.
32. A method for inhibiting the transcription modulating activity of P/CAF in a subject, comprising administering to the subject a transcription modulating activity inhibiting amount of a substance in a pharmaceutically acceptable carrier.
33. The method of claim 32, wherein the substance can inhibit the transcription modulating activity of P/CAF by preventing the binding of P/CAF to p300/CBP.
34. A method for stimulating the transcription modulating activity of P/CAF in a subject, comprising administering to the subject a transcription modulating activity stimulating amount of a substance in a pharmaceutically acceptable carrier.
- 35. The method of claim 34, wherein the substance can stimulate the transcription modulating activity of P/CAF by promoting the binding of P/CAF to p300/CBP.
36. The method of claim 34, wherein the substance can stimulate the transcription modulating activity of P/CAF by stimulating the histone acetyltransferase activity of P/CAF.
37. A method for inhibiting the histone acetyltransferase activity of p300/CBP in a subject, comprising administering to the subject a histone acetyltransferase activity inhibiting amount of a substance in a pharmaceutically acceptable carrier.

38. The method of claim 37, wherein the substance can inhibit the transcription modulating activity of p300/CBP by preventing the binding of a DNA-binding transcription factor to p300/CBP.

39. The method of claim 38, wherein the DNA-binding transcription factor is selected from the group consisting of a nuclear hormone receptor, CREB, c-Jun/v-Jun, c-Myb/v-Myb, YY1, Sap-1a, c-Fos, MyoD and SRC-1.

40. The method of claim 37, wherein the substance is an antibody which binds p300/CBP.

41. A method for stimulating the histone acetyltransferase activity of p300/CBP in a subject, comprising administering to the subject a histone acetyltransferase activity stimulating amount of a substance in a pharmaceutically acceptable carrier.

42. The method of claim 41, wherein the substance can stimulate the histone acetyltransferase activity of p300/CBP by promoting the binding of a DNA-binding transcription factor to p300/CBP.

1/3

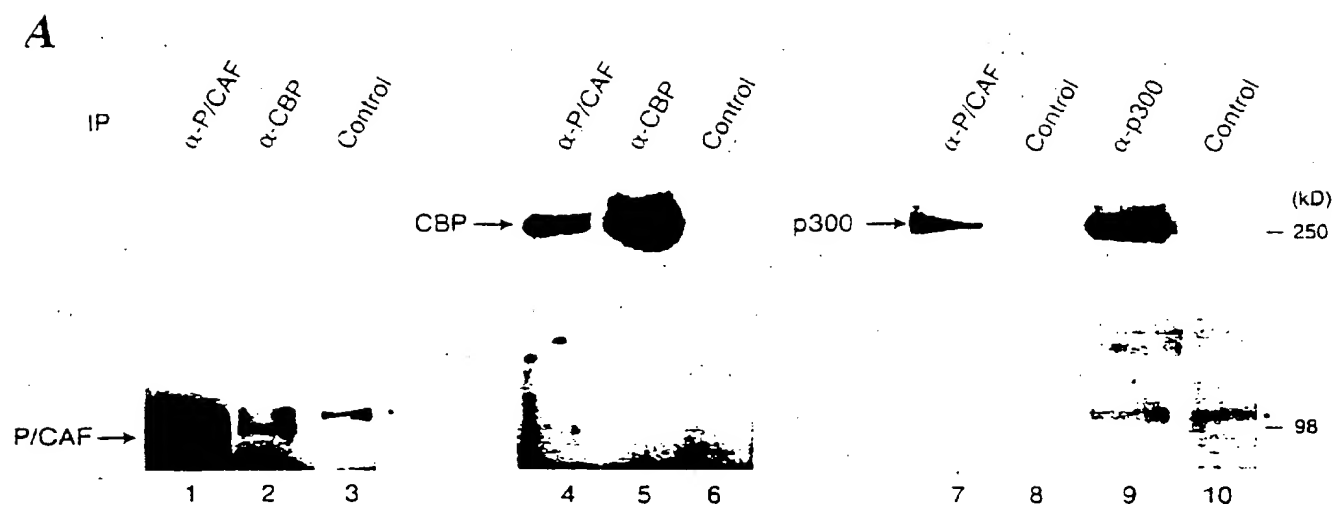
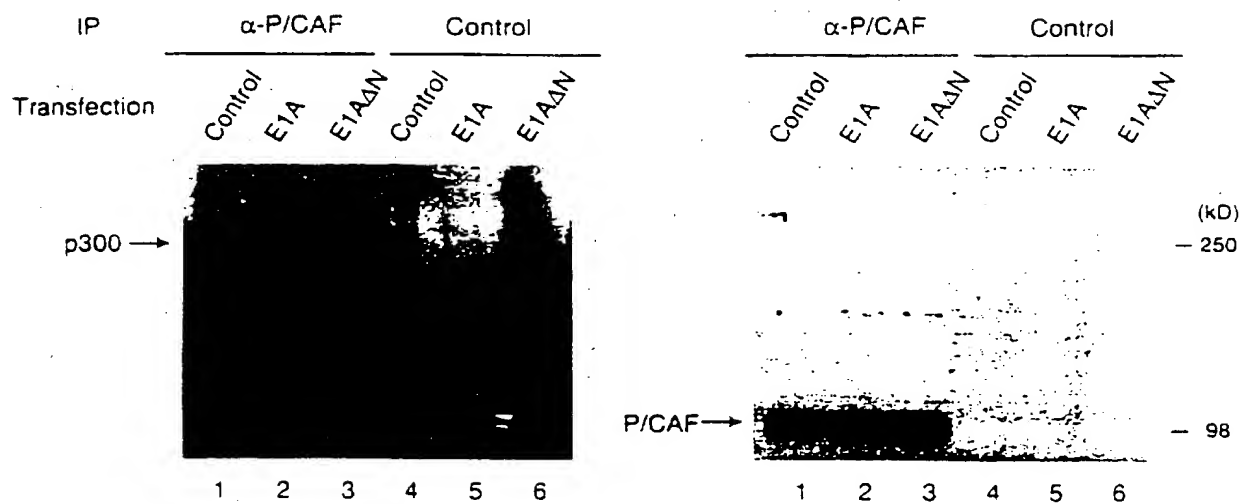
**B**

Fig. 1

2/3

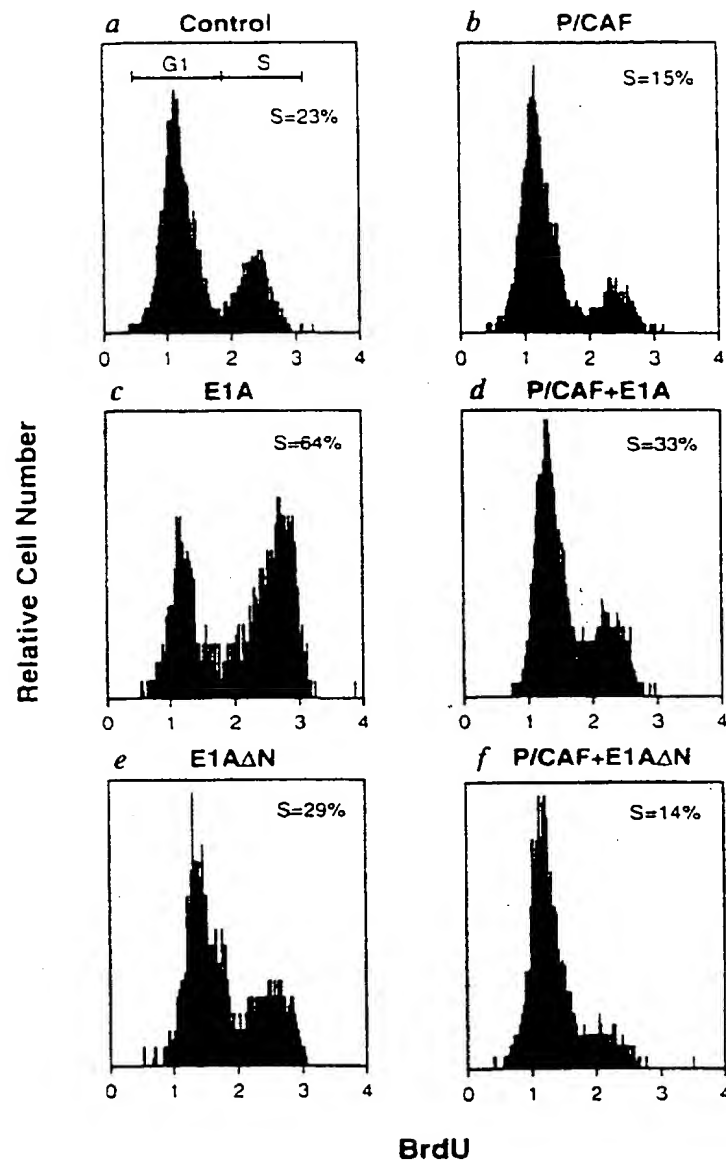


Fig. 2

3/3

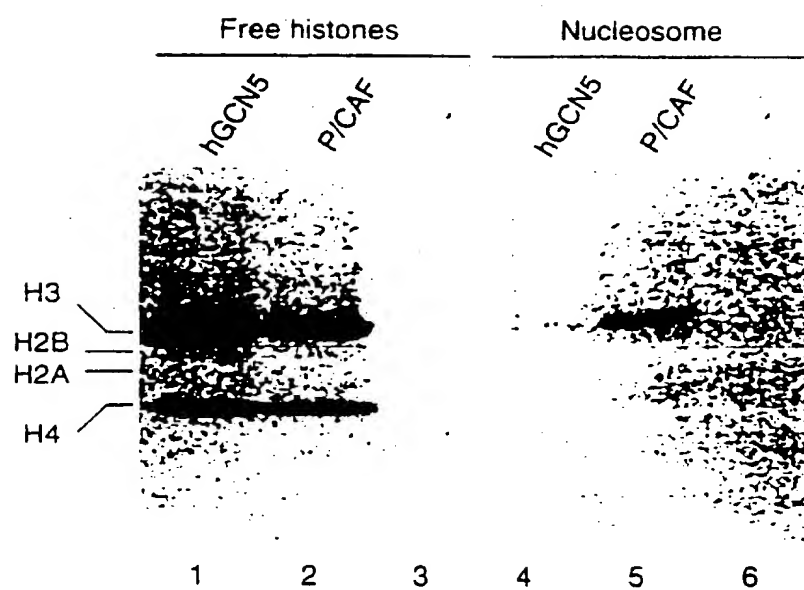


Fig. 3

**THIS PAGE BLANK (USPTO)**



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> :C12N 15/12, C07K 14/47, G01N 33/50,  
A61K 38/17

A3

(11) International Publication Number:

WO 98/03652

(43) International Publication Date:

29 January 1998 (29.01.98)

(21) International Application Number: PCT/US97/12877

(22) International Filing Date: 23 July 1997 (23.07.97)

(30) Priority Data:

60/022,273

23 July 1996 (23.07.96)

US

(60) Parent Application or Grant

(63) Related by Continuation

US

Filed on

60/022,273 (CIP)

23 July 1996 (23.07.96)

(71) Applicant (for all designated States except US): THE GOVERNMENT OF THE UNITED STATES OF AMERICA, represented by THE SECRETARY, DEPARTMENT OF HEALTH AND HUMAN SERVICES [US/US]; National Institutes of Health, Office of Technology Transfer Suite 325, 6011 Executive Boulevard, Rockville, MD 20852 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): NAKATANI, Yoshihiro [JP/US]; 4624 Edgefield Road, Bethesda, MD 20814 (US). HOWARD, Bruce, H. [US/US]; 8715 Fallen Oak Drive, Bethesda, MD 20817 (US).

(74) Agents: MILLER, Mary, L. et al.; Needle & Rosenberg, Suite 1200, 127 Peachtree Street, N.E., Atlanta, GA 30303 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

## Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

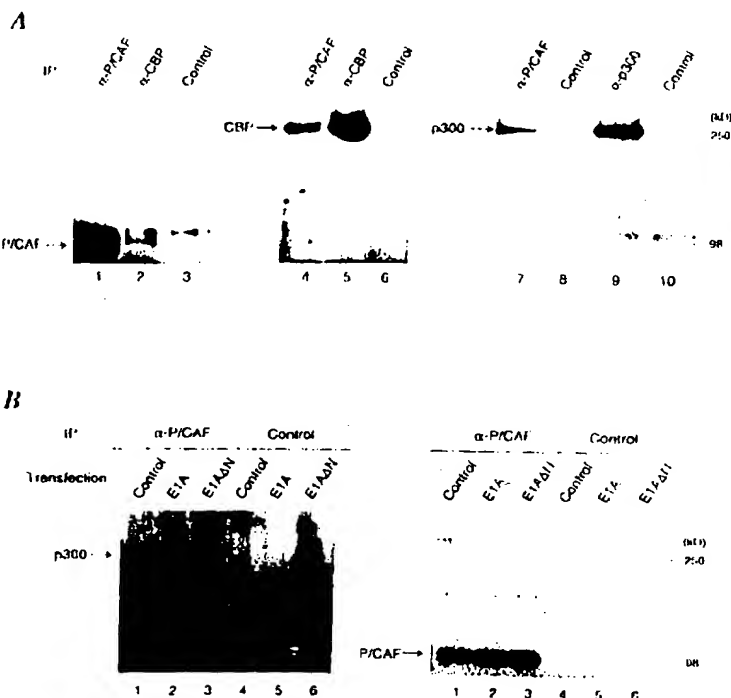
(88) Date of publication of the international search report:

26 February 1998 (26.02.98)

(54) Title: P300/CBP-ASSOCIATED TRANSCRIPTIONAL CO-FACTOR P/CAF AND USES THEREOF

## (57) Abstract

The present invention provides a purified protein designated P/CAF having a molecular weight of about 93,000 daltons as determined by sodium dodecyl sulfate polyacrylamide gel electrophoresis under reducing conditions and which acetylates histones and which also binds to the p300/CBP cellular protein. The present invention further provides a nucleic acid encoding the P/CAF protein as well as a vector containing the nucleic acid and a host for the vector. A purified antibody which specifically binds the P/CAF protein is also provided. Also provided are methods of screening for compounds that inhibit or stimulate the transcription modulating and histone acetyltransferase activity of P/CAF and p300/CBP.



**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

# INTERNATIONAL SEARCH REPORT

National Application No.

PCT/US 97/12877

**A. CLASSIFICATION OF SUBJECT MATTER**

IPC 6 C12N15/12 C07K14/47 G01N33/50 A61K38/17

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EMBL EST, Accession number N39522 Sequence no yv27b08.s1 Homo sapiens cDNA clone 243927 3' 25 January 1996 XP002050402 see the whole document	1
A	--- GEORGAKOPOULOS, T. & THIREOS, G.: EMBO JOURNAL., vol. 11, 1992, EYNSHAM, OXFORD GB, pages 4145-4152, XP002050399 see the whole document --- -/--	1

☒ Further documents are listed in the continuation of box C.

☐ Patent family members are listed in annex.

**\* Special categories of cited documents:**

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

\*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

\*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

\*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

\*Z\* document member of the same patent family

Date of the actual completion of the international search

17 December 1997

Date of mailing of the international search report

14. 01. 98

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl.  
Fax: (+31-70) 340-3016

Authorized officer

Chambonnet, F

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 97/12877

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	YANG, X.Y. ET AL.: "A p300-CBP-associated factor that competes with the adenoviral oncoprotein E1A" NATURE., vol. 382, no. 8589, 25 July 1996, LONDON GB, pages 319-324, XP002050400 see the whole document ---	1
P,X	OGRYZKO, V.V. ET AL.: "The transcriptional coactivators p300 and CBP are histone acetyltransferases" CELL, vol. 87, no. 5, November 1996, NA US, pages 953-959, XP002050401 see the whole document ---	1
P,X	EMBL EST, Accession number U57316, Sequence reference human GCN5 (hGCN5) complete cds. 26 august 1996 XP002050403 see the whole document -----	1

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US 97/12877

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:  
see FURTHER INFORMATION sheet PCT/ISA/210
2. ☐ Claims Nos.:  
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Remark : Although claims 32 to 42 are directed to a method of treatment of the human/animal body , the search has been carried out and based on the alleged effects of the composition.



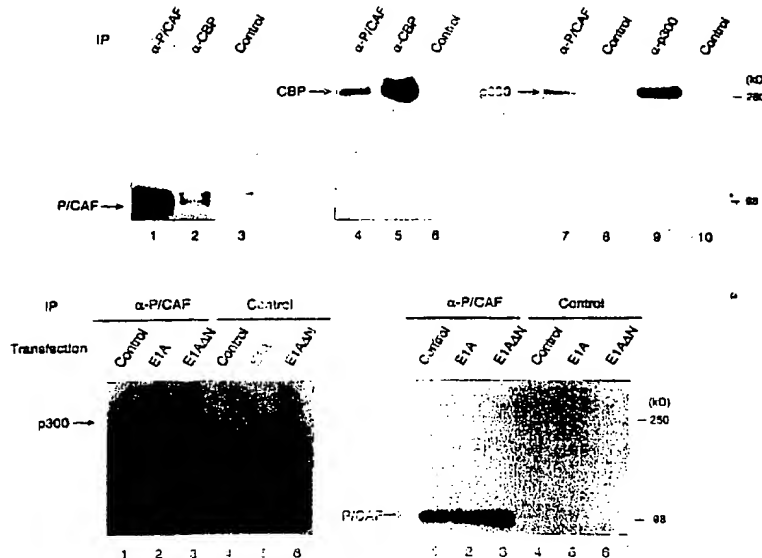
## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> : <b>C12N 15/12, C07K 14/47, G01N 33/50, A61K 38/17</b>		(11) International Publication Number: <b>WO 98/03652</b>
<b>A3</b>		(43) International Publication Date: 29 January 1998 (29.01.98)
(21) International Application Number: <b>PCT/US97/12877</b>		(74) Agents: MILLER, Mary, L. et al.; Needle & Rosenberg, Suite 1200, 127 Peachtree Street, N.E., Atlanta, GA 30303 (US).
(22) International Filing Date: 23 July 1997 (23.07.97)		
(30) Priority Data: 60/022,273 23 July 1996 (23.07.96) US		
(60) Parent Application or Grant (63) Related by Continuation US 60/022,273 (CIP) Filed on 23 July 1996 (23.07.96)		
(71) Applicant (for all designated States except US): THE GOVERNMENT OF THE UNITED STATES OF AMERICA, represented by THE SECRETARY, DEPARTMENT OF HEALTH AND HUMAN SERVICES [US/US]; National Institutes of Health, Office of Technology Transfer Suite 325, 6011 Executive Boulevard, Rockville, MD 20852 (US).		
(72) Inventors; and (75) Inventors/Applicants (for US only): NAKATANI, Yoshihiro [JP/US]; 4624 Edgefield Road, Bethesda, MD 20814 (US). HOWARD, Bruce, H. [US/US]; 8715 Fallen Oak Drive, Bethesda, MD 20817 (US).		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).
		Published With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.
		(88) Date of publication of the international search report: 26 February 1998 (26.02.98)

(54) Title: P300/CBP-ASSOCIATED TRANSCRIPTIONAL CO-FACTOR P/CAF AND USES THEREOF

## (57) Abstract

The present invention provides a purified protein designated P/CAF having a molecular weight of about 93,000 daltons as determined by sodium dodecyl sulfate polyacrylamide gel electrophoresis under reducing conditions and which acetylates histones and which also binds to the p300/CBP cellular protein. The present invention further provides a nucleic acid encoding the P/CAF protein as well as a vector containing the nucleic acid and a host for the vector. A purified antibody which specifically binds the P/CAF protein is also provided. Also provided are methods of screening for compounds that inhibit or stimulate the transcription modulating and histone acetyltransferase activity of P/CAF and p300/CBP.



**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

## P300/CBP-ASSOCIATED TRANSCRIPTIONAL CO-FACTOR P/CAF AND USES THEREOF

**BACKGROUND OF THE INVENTION**

5

**Field of the Invention**

The present invention provides a transcriptional co-factor, p300/CBP-associated factor (P/CAF), which modulates transcription through binding to the cellular transcription co-factors p300 and CBP and through acetylation of histones. Also  
10 provided are methods for screening for the presence of P/CAF and for substances which alter the transcription modulating effect and growth regulatory activity of P/CAF.

**Background Art**

Cellular proteins p300 and CBP are global transcriptional coactivators that are  
15 involved in the regulation of various DNA-binding transcriptional factors (Janknecht and Hunter, 1996). Recently, p300 was found to be very closely related to CBP, a factor that binds selectively to the protein kinase A-phosphorylated form of CREB (3-5). Cellular factors p300 and CBP exhibit strong amino acid sequence similarity and share the capacity to bind both CREB and E1A (6-8). Although neither p300 nor CBP by  
20 itself binds to DNA, each can be recruited to promoter elements via interaction with sequence-specific activators and functions to be a transcriptional adaptor. For simplicity, p300 and CBP will be termed p300/CBP in the context of discussing their shared functional properties.

25 p300/CBP is a large protein consisting of over 2,400 amino acids, known to interact with a variety of DNA-binding transcriptional factors including nuclear hormone receptors (13,57), CREB (3,4, 7), c-Jun/v-Jun (9,11), YY1 (10), c-Myb/v-Myb (12,58), Sap-1a (59), c-Fos (11) and MyoD (60). DNA-binding factors recruit p300/CBP not only by direct but also indirect interactions through cofactors; for example, nuclear  
30 hormone receptors recruit p300/CBP directly as well as through indirect interactions, via SRC-1, which stimulates transcription by binding to various nuclear hormone receptors (13,61).

The transforming proteins encoded by adenovirus and several other small DNA tumor viruses disturb host cell growth control by interacting with cellular factors that normally function to repress cell proliferation. One of the most intensively studied of these viral proteins, the product of the adenovirus E1A gene, is itself sufficient for transformation (1). E1A transforming activity resides in two distinct domains, the targets of which include p300/CBP and products of the retinoblastoma (RB) susceptibility gene family (1,2). Interactions of E1A with p300/CBP and RB are thought to influence functionally distinct growth regulatory pathways, allowing the two domains to contribute additively to transformation (1).

10

The paradigm for how E1A and functionally related viral proteins perturb cell growth regulation derives in large part from studies on their interactions with RB (1,2). The molecular function of E1A is based on its capacity to interfere with cellular protein-protein interactions. Since both E1A and various cellular targets bind to a site in RB termed the pocket domain (2), E1A can competitively disrupt the complex formation between RB and its cellular targets.

15

The second cellular factor implicated in E1A-dependent transformation, p300, is believed to inhibit G0/G1 exit, to activate certain enhancers, and to stimulate differentiation (1,2). E1A inhibits the p300/CBP-mediated transcriptional activation of many promoters (14). In one case that has been examined, the complex of p300 and YY1, E1A inhibits transcription without disrupting the complex (10).

20

The present invention provides a cellular protein designated P/CAF which binds to p300/CBP and plays an important role in both transcription and cell cycle regulation associated with a histone acetyltransferase activity. The present invention also provides a histone acetyltransferase activity in the p300/CBP cellular protein, thus providing targets for modulating transcription and cell cycle regulation in cells.

25

30

## SUMMARY OF THE INVENTION

The present invention provides a purified protein designated P/CAF having a molecular weight of about 93,000 daltons as determined by sodium dodecyl sulfate polyacrylamide gel electrophoresis under reducing conditions and which acetylates histones and which also binds to the p300/CBP cellular protein.

The present invention further provides a nucleic acid encoding the P/CAF protein as well as a vector containing the nucleic acid and a host for the vector. A purified antibody which specifically binds the P/CAF protein is also provided.

In addition, also provided is a bioassay for screening substances for the ability to inhibit the transcription modulating activity of P/CAF and/or histone acetyltransferase activity, comprising contacting the substance with a system in which histone acetylation by P/CAF can be determined; determining the amount of histone acetylation by P/CAF in the presence of the substance; and comparing the amount of histone acetylation by P/CAF in the presence of the substance with the amount of histone acetylation by P/CAF in the absence of the substance, a decreased amount of histone acetylation by P/CAF in the presence of the substance indicating a substance that can inhibit the transcription modulating activity and/or histone acetyltransferase activity of P/CAF.

Furthermore, the present invention provides a bioassay for screening substances for the ability to inhibit the transcription modulating activity and/or histone acetyltransferase activity of P/CAF comprising contacting the substance with a system in which the p300 binding of P/CAF can be determined; determining the amount of p300 binding of P/CAF in the presence of the substance; and comparing the amount of p300 binding of P/CAF in the presence of the substance with the amount of p300 binding of P/CAF in the absence of the substance, a decreased amount of p300 binding of P/CAF in the presence of the substance indicating a substance that can inhibit the transcription modulating activity and/or histone acetyltransferase activity of P/CAF.

Also provided is a method for determining the amount of P/CAF in a biological sample comprising contacting the biological sample with a polypeptide comprising the amino acid sequence of SEQ ID NO:3 under conditions whereby a P/CAF/p300 complex can be formed; and determining the amount of the P/CAF/p300 complex, the amount of the complex indicating the amount of P/CAF in the sample.

The present invention additionally provides a method for determining the amount of P/CAF in a biological sample comprising contacting the biological sample with an antibody which specifically binds P/CAF under conditions whereby a P/CAF/antibody complex can be formed; and determining the amount of the P/CAF/antibody complex, the amount of the complex indicating the amount of P/CAF in the sample.

Also provided herein is an assay for screening substances for the ability to inhibit or stimulate the histone acetyltransferase activity of P/CAF, comprising: contacting the substance with a system in which histone acetylation by P/CAF can be determined; determining the amount of histone acetylation by P/CAF in the presence of the substance; and comparing the amount of histone acetylation by P/CAF in the presence of the substance with the amount of histone acetylation by P/CAF in the absence of the substance, a decreased or increased amount of histone acetylation by P/CAF in the presence of the substance indicating a substance that can inhibit or stimulate, respectively, the histone acetyltransferase activity of P/CAF.

The present invention further provides an assay for screening substances for the ability to inhibit binding of P/CAF to p300/CBP comprising: contacting the substance with a system in which the P/CAF binding of p300/CBP can be determined; determining the amount of P/CAF binding of p300/CBP in the presence of the substance; and comparing the amount of binding of P/CAF to p300/CBP in the presence of the substance with the amount of binding of P/CAF to p300/CBP in the absence of the substance, a decreased amount of binding of P/CAF to p300/CBP in the presence of the substance indicating a substance that can inhibit the ability to inhibit binding of P/CAF to p300/CBP.

In addition, an assay is provided for screening substances for the ability to inhibit or stimulate the histone acetyltransferase activity of p300/CBP, comprising: contacting the substance with a system in which histone acetylation by p300/CBP can be  
5 determined; determining the amount of histone acetylation by p300/CBP in the presence of the substance; and comparing the amount of histone acetylation by p300/CBP in the presence of the substance with the amount of histone acetylation by p300/CBP in the absence of the substance, a decreased or increased amount of histone acetylation by p300/CBP in the presence of the substance indicating a substance that can inhibit or  
10 stimulate, respectively, the histone acetyltransferase activity of p300/CBP.

Furthermore, the present invention provides an assay for screening substances for the ability to inhibit binding of a DNA-binding transcription factor to p300/CBP comprising: contacting the substance with a system in which the DNA-binding  
15 transcription factor binding of p300/CBP can be determined; determining the amount of DNA-binding transcription factor binding of p300/CBP in the presence of the substance; and comparing the amount of binding of DNA-binding transcription factor to p300/CBP in the presence of the substance with the amount of binding of DNA-binding transcription factor to p300/CBP in the absence of the substance, a decreased amount of  
20 binding of DNA-binding transcription factor to p300/CBP in the presence of the substance indicating a substance that can inhibit the ability to inhibit binding of DNA-binding transcription factor to p300/CBP.

A method is also provided for inhibiting the transcription modulating activity of  
25 P/CAF in a subject, comprising administering to the subject a transcription modulating activity inhibiting amount of a substance in a pharmaceutically acceptable carrier.

Also provided in the present invention is a method for stimulating the transcription modulating activity of P/CAF in a subject, comprising administering to the  
30 subject a transcription modulating activity stimulating amount of a substance in a pharmaceutically acceptable carrier.

Furthermore, the present invention provides a method for inhibiting the histone acetyltransferase activity of p300/CBP in a subject, comprising administering to the subject a histone acetyltransferase activity inhibiting amount of a substance in a  
5 pharmaceutically acceptable carrier.

Finally, the present invention additionally provides a method for stimulating the histone acetyltransferase activity of p300/CBP in a subject, comprising administering to the subject a histone acetyltransferase activity stimulating amount of a substance in a  
10 pharmaceutically acceptable carrier.

### BRIEF DESCRIPTION OF THE FIGURES

Figs. 1A-B. Fig 1A: P/CAF-p300/CBP interaction *in vivo*. Cell extract was  
15 immunoprecipitated with rabbit anti-P/CAF (lanes 1, 4, and 7), rabbit anti-CBP (lanes 2 and 5), and mouse anti-p300 (lane 9) antibodies. For controls, cell extract was precipitated with rabbit control IgG (lanes 3, 6, and 8) or mouse anti-HA monoclonal antibody (lane 10). The precipitates were analyzed by immunoblotting with anti-P/CAF (lanes 1-3), anti-CBP (lanes 4-6), and anti-p300 (lanes 7-10) antibodies. The positions  
20 of non-specific bands are indicated by asterisks. Fig. 1B: E1A inhibits the P/CAF-p300 interaction *in vivo*. Osteosarcoma cells were transfected with either control vector (lanes 1 and 4) or E1A- (lanes 2 and 5) or E1AΔN- (lanes 3 and 6) expression vectors. Extract from the transfected subpopulation was immunoprecipitated with anti-P/CAF (lanes 1-3) or control (lanes 4-6) IgG. The precipitates were analyzed by  
25 immunoblotting with anti-p300 and anti-P/CAF.

Figs. 2A-F. P/CAF and E1A mediate antagonistic effects on cell cycle progression. HeLa cells (ATCC accession number CCL 2) were transfected by electroporation with 7  $\mu$ g of P/CAF-expression plasmid and/or 3  $\mu$ g of the full-length or  
30 the N-terminally deleted ( $\Delta$ 2-36) E1A 12S-expression plasmid as indicated in the figure. These plasmids were constructed by subcloning FLAG-P/CAF and E1A cDNAs into

pCX (34) and pcDNAI (Invitrogen), respectively. All samples, in addition, contained 1  $\mu$ g of sorting plasmid (pCMV-IL2R) (31) and carrier plasmid (pCX) to normalize the total amount of DNA to 11  $\mu$ g. After transfection, cells were incubated in Dulbecco's modified Eagle's medium with 10% fetal bovine calf serum for 12 hours and subsequently labeled in medium containing 10  $\mu$ M bromo-deoxyuridine (BrdU) for 30 min. Subsequently, the transfected subpopulation was purified by magnetic affinity cell sorting and nuclei were analyzed by dual parameter flow cytometry as described (32). Histograms show percentages of cells in G1 and S phases. Abscissa values represent fluorescence intensity of bound anti-BrdU antibodies in log scale.

10

Fig. 3. Histone acetyltransferase activity of P/CAF. Activity of hGCN5 (lanes 1 and 4) and P/CAF (lanes 2 and 5) that acetylates free histones (lanes 1-3) or histones in the nucleosome core particle (35) (lanes 4-6) was measured as described (36). Each reaction contains 0.3 pmol of affinity purified FLAG-hGCN5 or FLAG-P/CAF, 4 pmol of the histone octamer or the nucleosome core particle and 10 pmol of [1- $^{14}$ C]acetyl-CoA. Note that the histone octamer dissociates into dimers or tetramers under assay conditions. Acetylated histones were detected by autoradiography after separation by SDS-PAGE. The bands corresponding to acetylated histones H3 and H4 are indicated by arrows.

20

## DETAILED DESCRIPTION OF THE INVENTION

As used in the specification and in the claims, "a" can mean one or more, depending upon the context in which it is used.

25

### P/CAF protein and fragments

The present invention provides a purified protein designated P/CAF having a molecular weight of about 93,000 daltons as determined by sodium dodecyl sulfate polyacrylamide gel electrophoresis under reducing conditions and which acetylates histones. The P/CAF protein can also bind to the amino acid region of SEQ ID NO:3

30

(amino acid (aa) residues 1753 - 1966) of the cellular transcriptional factor, p300 (which has the complete amino acid sequence of SEQ ID NO:6 and the nucleotide sequence of SEQ ID NO:12), and the amino acid region of SEQ ID NO:6 (amino acid residues 1805 - 1854) of the cellular transcriptional factor, CBP (which has the complete amino acid sequence of SEQ ID NO:7 and the nucleotide sequence of SEQ ID NO:13). The P/CAF protein can be defined by any one or more of the typically used parameters. Examples of these parameters include, but are not limited to molecular weight (calculated or empirically determined), isoelectric focusing point, specific epitope(s), complete amino acid sequence, sequence of a specific region (e.g., N-terminus) of the amino acid sequence and the like.

For example, The P/CAF protein can consist of the amino acid sequence of SEQ ID NO:1 or the P/CAF protein can comprise the amino acid sequence of SEQ ID NO:2 which represents the carboxy terminal end of the P/CAF protein and contains the histone acetyltransferase activity, or the amino acid sequence of SEQ ID NO:4, which represents the amino terminal end of the P/CAF protein, containing the binding site for p300/CBP. Because the amino-terminal region is specific for P/CAF it can be used to define and identify P/CAF.

As used herein, "purified" refers to a protein (polypeptide, peptide, etc.) that is sufficiently free of contaminants or cell components with which it normally occurs to distinguish it from the contaminants or other components of its natural environment. The purified protein need not be homogeneous, but must be sufficiently free of contaminants to be useful in a clinical or research setting, for example, in an assay for detecting antibodies to the protein. Greater levels of purity can be obtained using methods derived from well known protocols. Specific methods for purifying P/CAF proteins are known in the art.

As will be appreciated by those skilled in the art, the invention also includes those P/CAF polypeptides having slight variations in amino acid sequence which yield polypeptides equivalent to the P/CAF protein defined herein. Such variations may arise

naturally as allelic variations (*e.g.*, due to genetic polymorphism) or may be produced by human intervention (*e.g.*, by mutagenesis of cloned DNA sequences), such as induced point, deletion, insertion and substitution mutants. Minor changes in amino acid sequence are generally preferred, such as conservative amino acid replacements, small internal deletions or insertions, and additions or deletions at the ends of the molecules. Substitutions may be designed based on, for example, the model of Dayhoff, *et al.* (37). These modifications can result in changes in the amino acid sequence, provide silent mutations, modify a restriction site, or provide other specific mutations.

- 10            Modifications to any of the P/CAF proteins or fragments can be made, while preserving the specificity and activity (function) of the native protein or fragment thereof. As used herein, "native" describes a protein that occurs in nature. The modifications contemplated herein can be conservative amino acid substitutions, for example, the substitution of a basic amino acid for a different basic amino acid.
- 15            Modifications can also include creation of fusion proteins with epitope tags or known recombinant proteins or genes encoding them created by subcloning into commercial or non-commercial vectors (*e.g.*, polyhistidine tags, flag tags, myc tag, glutathione-S-transferase [GST] fusion protein, *xyIE* fusion reporter construct). Furthermore, the modifications can be such as do not affect the function of the protein or the way the
- 20            protein accomplishes that function (*e.g.*, its secondary structure or the ultimate result of the protein's activity). These products are equivalent to the P/CAF protein. The means for determining the function, way and result parameters are well known.

- Having provided an example of a purified P/CAF protein, the invention also
- 25            enables the purification of P/CAF homologs from other species and allelic variants from individuals within a species. For example, an antibody raised against the exemplary human P/CAF protein can be used routinely to screen preparations from different humans for allelic variants of the P/CAF protein that react with the P/CAF protein-specific antibody. Similarly, an antibody raised against an epitope, for example, from a
- 30            conserved amino acid region of the human P/CAF protein can be used to routinely screen for homologs of the P/CAF protein in other species. A P/CAF protein can be

routinely identified in and obtained from other species and from individuals within a species using the methods taught herein and others known in the art. For example, given the present sequence, the DNA encoding a conserved amino acid sequence can be used to probe genomic DNA or DNA libraries of an organism to predictably obtain the P/CAF gene for that organism. The gene can then be cloned and expressed as the P/CAF protein and purified according to any of a number of routine, predictable methods. An example of the routine protein purification methods available in the art can be found in Pei et al. (38).

10 A purified polypeptide fragment of the P/CAF protein is also provided. The term "fragment" as used herein regarding a P/CAF protein, means a molecule of at least five contiguous amino acids of P/CAF protein that has at least one function shared by P/CAF protein or a region thereof. These functions can include antigenicity, binding capacity, acetyltransferase activity and structural roles, among others. The P/CAF  
15 fragment can be specific for a recited source. As used herein to describe an amino acid sequence (protein, polypeptide, peptide, etc.), "specific" means that the amino acid sequence is not found identically in any other source. The determination of specificity is made routine by the availability of computerized amino acid sequence databases and sequence comparison programs, wherein an amino acid sequence of almost any length  
20 can be quickly and reliably checked for the existence of identical sequences. If an identical sequence is not found, the protein is "specific" for the recited source. For example, a P/CAF fragment can be species-specific (e.g., found in the P/CAF protein of humans, but not of other species).

25 A fragment of the P/CAF protein having histone acetyltransferase activity can consist of the amino acid sequence of SEQ ID NO:2. A fragment of the P/CAF protein which binds to the amino acid sequence of SEQ ID NO:3 on p300 and the amino acid sequence of SEQ ID NO:9 on CBP can consist of the amino acid sequence of SEQ ID NO:4. To the extent that these fragments are specific for P/CAF, they can be used to  
30 identify and define P/CAF.

An antigenic fragment of P/CAF protein is provided. An antigenic fragment has an amino acid sequence of at least about five consecutive amino acids of a P/CAF protein amino acid sequence and binds an antibody or elicits an immune response in an animal. An antigenic fragment can be selected by applying the routine technique of epitope mapping to P/CAF protein to determine the regions of the proteins that contain epitopes reactive with antibodies or are capable of eliciting an immune response in an animal. Once the epitope is selected, an antigenic polypeptide containing the epitope can be synthesized directly, or produced recombinantly by cloning nucleic acids encoding the antigenic polypeptide in an expression system, according to standard methods.

Alternatively, an antigenic fragment of the antigen can be isolated from the whole P/CAF protein or a larger fragment of the P/CAF protein by chemical or mechanical disruption. Fragments can also be randomly chosen from a known P/CAF protein sequence and synthesized. The purified fragments thus obtained can be tested to determine their antigenicity and specificity by routine methods.

#### **Nucleic Acids Encoding P/CAF Protein**

An isolated nucleic acid that encodes a P/CAF protein is also provided. As used herein, the term "isolated" means a nucleic acid separated or substantially free from at least some of the other components of the naturally occurring organism, for example, the cell structural components commonly found associated with nucleic acids in a cellular environment and/or other nucleic acids. The isolation of nucleic acids can therefore be accomplished by techniques such as cell lysis followed by phenol plus chloroform extraction, followed by ethanol precipitation of the nucleic acids (39). It is not contemplated that the isolated nucleic acids are necessarily totally free of all non-nucleic acid components or all other nucleic acids, but that the isolated nucleic acids are isolated to a degree of purification to be useful in clinical, diagnostic, experimental, or other procedures such as, for example, gel electrophoresis, Southern, Northern or dot blot hybridization, or polymerase chain reaction (PCR).

A skilled artisan in the field will readily appreciate that there are a multitude of procedures which may be used to isolate the nucleic acids prior to their use in other procedures. These include, but are not limited to, lysis of the cell followed by gel filtration or anion exchange chromatography, binding DNA to silica in the form of glass  
5 beads, filters or diatoms in the presence of high concentrations of chaotropic salts, or ethanol precipitation of the nucleic acids.

The nucleic acids of the present invention can include positive and negative strand RNA as well as DNA and can include genomic and subgenomic nucleic acids  
10 found in the naturally occurring organism. The nucleic acids contemplated by the present invention include double stranded and single stranded DNA of the genome, complementary positive stranded cRNA and mRNA, and complementary cDNA produced therefrom and any nucleic acid which can selectively or specifically hybridize to the isolated nucleic acids provided herein. Stringent conditions (further described  
15 below) are used to distinguish selectively or specifically hybridizing nucleic acids from non-selectively and non-specifically hybridizing nucleic acids.

An isolated nucleic acid that encodes a P/CAF protein can be species-specific (i.e., does not encode the P/CAF protein of other species and does not occur in other  
20 species). Examples of the nucleic acids contemplated herein include the nucleic acid of SEQ ID NO:10 as well as the nucleic acids that encode each of the P/CAF proteins or fragments thereof described herein. P/CAF proteins and protein fragments can be routinely obtained as described herein and their structure (sequence) determined by routine means including the methods as used herein.

25

P/CAF protein-encoding nucleic acids can be isolated from an organism in which they are normally found (e.g., humans), using any of the routine techniques. For example, a genomic DNA or cDNA library can be constructed and screened for the presence of the nucleic acid of interest using one of the present P/CAF protein-encoding  
30 nucleic acids as a probe. Methods of constructing and screening such libraries are well known in the art and kits for performing the construction and screening steps are

commercially available (for example, Stratagene Cloning Systems, La Jolla, CA). Once isolated, the nucleic acid can be directly cloned into an appropriate vector, or if necessary, be modified to facilitate the subsequent cloning steps. Such modification steps are routine, an example of which is the addition of oligonucleotide linkers, which contain restriction sites, to the termini of the nucleic acid (See, for example, ref. 39).

P/CAF protein-encoding nucleic acids can also be synthesized. For example, a method of obtaining a DNA molecule encoding a specific P/CAF protein is to synthesize a recombinant DNA molecule which encodes the P/CAF protein. For example, nucleic acid synthesis procedures are routine in the art and oligonucleotides coding for a particular protein region are readily obtainable through automated DNA synthesis. A nucleic acid for one strand of a double-stranded molecule can be synthesized and hybridized to its complementary strand. One can design these oligonucleotides such that the resulting double-stranded molecule has either internal restriction sites or appropriate 5' or 3' overhangs at the termini for cloning into an appropriate vector.

Oligonucleotides complementary to or identical with the P/CAF protein-encoding nucleic acid sequence can be synthesized as primers for amplification reactions, such as PCR, or as probes to detect P/CAF protein encoding nucleic acids by various hybridization protocols (e.g., Northern blot; Southern blot; dot blot, colony screening, etc.).

Double-stranded molecules coding for relatively large proteins can readily be synthesized by first constructing several different double-stranded molecules that code for particular regions of the protein, followed by ligating these DNA molecules together. For example, Cunningham, et al. (40), have constructed a synthetic gene encoding the human growth hormone by first constructing overlapping and complementary synthetic oligonucleotides and ligating these fragments together. See also, Ferretti, et al. (41), wherein synthesis of a 1057 base pair synthetic bovine rhodopsin gene from synthetic oligonucleotides is disclosed.

By constructing a P/CAF protein-encoding nucleic acid in this manner, one skilled in the art can readily obtain any particular P/CAF protein with modifications at any particular position or positions. See also, U.S. Patent No. 5,503,995 which describes an enzyme template reaction method of making synthetic genes. Techniques  
5 such as this are routine in the art and are well documented. DNA encoding the P/CAF protein or P/CAF protein fragments can then be expressed *in vivo* or *in vitro*.

The nucleic acid encoding the P/CAF protein can be any nucleic acid that functionally encodes the P/CAF protein. To functionally encode the protein (i.e., allow  
10 the nucleic acid to be expressed), the nucleic acid can include, but is not limited to, expression control sequences, such as an origin of replication, a promoter, regions upstream or downstream of the promoter, such as enhancers that may regulate the transcriptional activity of the promoter, appropriate restriction sites to facilitate cloning of inserts adjacent to the promoter, antibiotic resistance genes or other markers which  
15 can serve to select for cells containing the vector or the vector containing the insert, and necessary information processing sites, such as ribosome binding sites, RNA splice sites, polyadenylation sites and transcription termination sequences as well as any other sequence which may facilitate the expression of the inserted nucleic acid.

20 Preferred expression control sequences are promoters derived from metallothionine genes, actin genes, immunoglobulin genes, CMV, SV40, adenovirus, bovine papilloma virus, etc. A nucleic acid encoding a P/CAF protein can readily be determined based upon the genetic code for the amino acid sequence of the P/CAF protein and many nucleic acid sequences will encode a P/CAF protein. Modifications in  
25 the nucleic acid sequence encoding the P/CAF protein are also contemplated. Modifications that can be useful are modifications to the sequences controlling expression of the P/CAF protein to make production of P/CAF protein inducible or repressible as controlled by the appropriate inducer or repressor. Such means are standard in the art (*see, e.g.,* ref. 39). The nucleic acids can be generated by means  
30 standard in the art, such as by recombinant nucleic acid techniques, as exemplified in the examples herein, and by synthetic nucleic acid synthesis or *in vitro* enzymatic synthesis.

After a nucleic acid encoding a particular P/CAF protein of interest, or a region of that nucleic acid, is constructed, modified, or isolated, that nucleic acid can then be cloned into an appropriate vector, which can direct the *in vivo* or *in vitro* synthesis of that wild-type and/or modified P/CAF protein. The vector is contemplated to have the necessary functional elements that direct and regulate transcription of the inserted nucleic acid, as described above. The vector containing the P/CAF nucleic acid or nucleic acid fragment can be in a host (e.g., cell or transgenic animal) for expressing the nucleic acid. The P/CAF protein or fragment thereof can thus be produced in a host system containing the expression vector and its functional activity as described herein can be demonstrated according to methods well known in the art.

There are numerous *E. coli* (*Escherichia coli*) expression vectors known to one of ordinary skill in the art useful for the expression of proteins. Other microbial hosts suitable for use include bacilli, such as *Bacillus subtilis*, and other enterobacteria, such as *Salmonella*, *Serratia*, as well as various *Pseudomonas* species. These prokaryotic hosts can support expression vectors which will typically contain expression control sequences compatible with the host cell (e.g., an origin of replication). In addition, any number of a variety of well-known promoters will be present, such as the lactose promoter system, a tryptophan (Trp) promoter system, a beta-lactamase promoter system, or a promoter system from phage lambda. The promoters will typically control expression, optionally with an operator sequence and have ribosome binding site sequences, for example, for initiating and completing transcription and translation. If necessary, an amino terminal methionine can be provided by insertion of a Met codon 5' and in-frame with the gene sequence. Also, the carboxy-terminal extension of the protein can be removed using standard oligonucleotide mutagenesis procedures.

Additionally, yeast expression can be used. There are several advantages to yeast expression systems. First, evidence exists that proteins produced in yeast secretion systems exhibit correct disulfide pairing. Second, post-translational glycosylation is efficiently carried out by yeast secretory systems. The *Saccharomyces cerevisiae* pre-pro-alpha-factor leader region (encoded by the *MF $\alpha$ -1* gene) is routinely used to direct

protein secretion from yeast (42). The leader region of pre-pro-alpha-factor contains a signal peptide and a pro-segment which includes a recognition sequence for a yeast protease encoded by the *KEX2* gene. This enzyme cleaves the precursor protein on the carboxyl side of a Lys-Arg dipeptide cleavage-signal sequence. The polypeptide coding sequence can be fused in-frame to the pre-pro-alpha-factor leader region. This construct is then put under the control of a strong transcription promoter, such as the alcohol dehydrogenase I promoter or a glycolytic promoter. The protein coding sequence is followed by a translation termination codon which is followed by transcription termination signals. Alternatively, the polypeptide encoding sequence of interest can be fused to a second protein coding sequence, such as Sj26 or  $\beta$ -galactosidase, used to facilitate purification of the resultant fusion protein by affinity chromatography. The insertion of protease cleavage sites to separate the components of the fusion protein is applicable to constructs used for expression in yeast.

Efficient post-translational glycosylation and expression of recombinant proteins can also be achieved in *Baculovirus* expression systems in insect cells.

Mammalian cells permit the expression of proteins in an environment that favors important post-translational modifications such as folding and cysteine pairing, addition of complex carbohydrate structures and secretion of active protein. Vectors useful for the expression of proteins in mammalian cells are characterized by insertion of the protein encoding sequence between a strong viral promoter and a polyadenylation signal. The vectors can contain genes conferring either gentamicin or methotrexate resistance for use as selectable markers. For example, the antigen and immunoreactive fragment coding sequence can be introduced into a Chinese hamster ovary (CHO) cell line using a methotrexate resistance-encoding vector. Presence of the vector RNA in transformed cells can be confirmed by Northern blot analysis and production of a cDNA or opposite strand RNA corresponding to the protein encoding sequence can be confirmed by Southern and Northern blot analysis, respectively. A number of other suitable host cell lines capable of secreting intact proteins have been developed in the art and include the CHO cell lines, HeLa cells, myeloma cell lines, Jurkat cells, and the like.

Expression vectors for these cells can include expression control sequences, as described above. The vectors containing the nucleic acid sequences of interest can be transferred into the host cell by well-known methods, which vary depending on the type of cell host. For example, calcium chloride transfection is commonly utilized for prokaryotic cells, 5 whereas calcium phosphate treatment or electroporation may be used for other cell hosts.

Alternative vectors for the expression of protein in mammalian cells, similar to those developed for the expression of human gamma-interferon, tissue plasminogen 10 activator, clotting Factor VIII, hepatitis B virus surface antigen, protease Nexin I, and eosinophil major basic protein, can be employed. Further, the vector can include CMV promoter sequences and a polyadenylation signal available for expression of inserted nucleic acid in mammalian cells (such as COS7).

15 The nucleic acid sequences can be expressed in hosts after the sequences have been positioned to ensure the functioning of an expression control sequence. These expression vectors are typically replicable in the host organisms either as episomes or as an integral part of the host chromosomal DNA. Commonly, expression vectors can contain selection markers, e.g., tetracycline resistance or hygromycin resistance, to 20 permit detection and/or selection of those cells transformed with the desired nucleic acid sequences (*see*, e.g., U.S. Patent 4,704,362).

The nucleic acids produced as described above can also be expressed in a host which is a non-human animal to create a transgenic animal, containing, in a germ or 25 somatic cell, a nucleic acid comprising the coding sequence for all or a portion of the P/CAF protein, as well as all of the other regulatory elements required for expression of the P/CAF protein-encoding sequence. The animal will express the P/CAF gene or portion thereof to produce the P/CAF protein or protein fragment and such expression can be detected by determination of a particular phenotype unique to the transgenic 30 animal expressing the transferred nucleic acid.

The nucleic acid can be the nucleic acid of SEQ ID NO:10, a nucleic acid having a nucleotide sequence which encodes the P/CAF protein, a nucleic acid having a nucleotide sequence which encodes the protein of SEQ ID NO:1, as well as the nucleic acids that encode the proteins comprising the fragments of SEQ ID NOS:2 and 4.

5

The nucleic acids of the invention can contain substitutions or deletions which provide a particular phenotype of interest. For example, various deletions or base substitutions can be introduced into the nucleic acid encoding the P/CAF protein for the purpose of studying the effects of these particular deletions or substitutions on the transcription modulation activity of the P/CAF protein. These effects can be monitored by observation of such characteristics as growth and development of the animal, the ability to develop tumors, survival rates and the like. The gene construct introduced into the animal cells to produce the transgenic animal can contain any of the regulatory elements described above to modulate expression of the foreign genes. As used herein, the term "phenotype" includes morphology, biochemical profiles, changes in tumor formation and other parameters that are affected by the presence of the P/CAF protein.

The transgenic animals of the invention can also be used in a method for determining the effectiveness of administering a nucleic acid encoding a functional P/CAF protein to a subject in need of a functional P/CAF protein. First, a nucleic acid encoding a nonfunctional P/CAF protein can be introduced into the animal's cells and expressed to yield a characteristic phenotype. Then, using standard gene therapy techniques, a nucleic acid encoding a functional P/CAF protein can be introduced into the animal's cells and the effects on the animal's phenotypic characteristics can be determined.

25

Having provided and taught how to obtain a nucleic acid that encodes a P/CAF protein, an isolated nucleic acid that encodes a fragment of P/CAF protein is also provided. The nucleic acid encoding the fragment can be obtained using any of the methods applicable to the nucleic acid encoding the entire P/CAF protein. The nucleic acid fragment can encode a species-specific P/CAF protein fragment (e.g., found in the

30

P/CAF protein of humans, but not in the P/CAF proteins of other species). Nucleic acids encoding species-specific fragments of P/CAF protein are themselves species-specific or allele-specific fragments of the P/CAF gene.

5           Examples of fragments of a nucleic acid encoding a fragment of the P/CAF protein can include the nucleic acid sequences which encode the amino acid sequences of the fragments of SEQ ID NOS:2 or 4. The same routine computer analyses used to select these examples of fragments can be routinely used to obtain others. Fragments of P/CAF-encoding nucleic acids can be primers for PCR or probes, which can be species-specific, gene-specific or allele-specific. P/CAF-encoding nucleic acid fragments can  
10           encode antigenic or immunogenic fragments of P/CAF protein that can be used in therapeutic assays or screening protocols. P/CAF gene fragments can encode fragments of P/CAF protein having histone acetylase activity and/or p300/CBP binding activity as described above, as well as other uses that may become apparent.

15

          An isolated nucleic acid of at least ten nucleotides that selectively hybridizes with the nucleic acid of SEQ ID NO:10 under selected conditions is provided. For example, the conditions can be PCR amplification conditions and the hybridizing nucleic acid can be a primer consisting of a specific fragment of the reference sequence or a nearly  
20           identical nucleic acid that hybridizes only to the exemplified P/CAF-encoding nucleic acid or allelic variants thereof.

          The invention provides an isolated nucleic acid that selectively hybridizes with the P/CAF-encoding nucleic acid sequence of SEQ ID NO:10 under stringent  
25           conditions. The hybridizing nucleic acid can be a probe that hybridizes only to the exemplified P/CAF-encoding nucleic acid sequence. Thus, the hybridizing nucleic acid can be a naturally occurring species-specific allelic variant of the exemplified P/CAF gene. The hybridizing nucleic acid can also include insubstantial base substitutions that do not prevent hybridization under the stated stringent conditions or affect either the  
30           function of the encoded protein, the way the protein accomplishes that function (e.g., its

secondary structure) or the ultimate result of the protein's activity. The means for determining these parameters are well known.

As used herein to describe nucleic acids, the term "selectively hybridizes" excludes the occasional randomly hybridizing nucleic acids as well as nucleic acids that encode other known homologs of the P/CAF protein. The selectively hybridizing nucleic acids of the invention can have at least 70%, 73%, 78%, 80%, 85%, 88%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% complementarity with the segment and strand of the sequence to which it hybridizes. This list is not intended to exclude percent complementarity values between these values. The nucleic acids can be at least 10, 15, 16, 17, 18, 20, 21, 23, 24, 25, 30, 35, 40, 50, 100, 150, 200, 300, 500, 550, 750, 900, 950, or 1000 nucleotides in length or any intervening length, depending on whether the nucleic acid is to be used as a primer, probe or for protein expression. The hybridizing nucleic acid can comprise a region of at least ten nucleotides (up to full length) that is completely complementary to a unique region of the nucleic acid to which it hybridizes.

The nucleic acid can be an alternative coding sequence for the P/CAF protein, or can be used as a probe or primer for detecting the presence of or obtaining the P/CAF protein. If used as primers, the invention provides compositions including at least two nucleic acids which selectively hybridize with different regions of the nucleic acid so as to amplify a desired region. Depending on the length of the probe or primer, it can range between 70% complementary bases and full complementarity and still hybridize under stringent conditions.

25

For example, for the purpose of obtaining or determining the presence of a nucleic acid encoding the P/CAF protein, the degree of complementarity between the hybridizing nucleic acid (probe or primer) and the sequence to which it hybridizes (P/CAF DNA in a sample) should be at least enough to exclude hybridization with a nucleic acid from another species. The invention provides examples of these nucleic acids of P/CAF, so that the degree of complementarity required to distinguish selectively

30

hybridizing from nonselectively hybridizing nucleic acids under stringent conditions can be clearly determined for each nucleic acid. It should also be clear that the hybridizing nucleic acids of the invention will not hybridize with nucleic acids encoding unrelated proteins (hybridization is selective) under stringent conditions.

5

"Stringent conditions" refers to the washing conditions used in a hybridization protocol. In general, the washing conditions should be a combination of temperature and salt concentration chosen so that the denaturation temperature is approximately 5-20°C below the calculated  $T_m$  of the nucleic acid hybrid under study. The temperature and salt conditions are readily determined empirically in preliminary experiments in which samples of reference DNA immobilized on filters are hybridized to the probe or protein encoding nucleic acid of interest and then washed under conditions of different stringencies. For example, the nucleic acid sequence of SEQ ID NO:10 was used as a specific radiolabeled probe for the detection of messenger RNA transcribed from the P/CAF gene by performing hybridizations under stringent conditions. The  $T_m$  of such an oligonucleotide can be estimated by allowing 2°C for each A or T nucleotide, and 4°C for each G or C. For example, an 18 nucleotide probe of 50% G+C would, therefore, have an approximate  $T_m$  of 54°C.

20 The invention provides an isolated nucleic acid that selectively hybridizes with the P/CAF gene shown in the sequence set forth as SEQ ID NO:10 under stringent conditions. The invention further provides an isolated nucleic acid complementary to the nucleotide sequence set forth in SEQ ID NO:10.

## 25 **Antibodies to the P/CAF protein**

A purified antibody and an antiserum containing polyclonal antibodies that specifically bind the P/CAF protein or antigenic fragment are also provided. The term "bind" means the well understood antigen/antibody binding as well as other nonrandom association with an antigen. "Specifically bind" as used herein describes an antibody or other ligand that does not cross react substantially with any antigen other than the one specified, in this case, an antigen of the P/CAF protein. Antibodies can be made as

30

described in Harlow and Lane (33). Briefly, purified P/CAF protein or an antigenic fragment thereof can be injected into an animal in an amount and in intervals sufficient to elicit a humoral immune response. Serum polyclonal antibodies can be purified directly, or spleen cells from the animal can be fused with an immortal cell line and screened for monoclonal antibody secretion, according to procedures well known in the art. Purified monospecific polyclonal antibodies that specifically bind the P/CAF antigen are also within the scope of the present invention. The antibodies of the present invention can bind the protein of claim 1, the protein of claim 2, the protein of claim 3 and/or the protein of claim 4, as well as any other proteins of the present invention.

10

A ligand that specifically binds the antigen is also contemplated. The ligand can be a fragment of an antibody, such as, for example, an Fab fragment which retains P/CAF binding activity, or a smaller molecule designed to bind an epitope of the P/CAF antigen. The antibody or ligand can be bound to a substrate or labeled with a detectable moiety or both bound and labeled. The detectable moieties contemplated within the compositions of the present invention include those listed above in the description of the diagnostic methods, including fluorescent, enzymatic and radioactive markers.

15

The antibody can be bound to a solid support substrate or conjugated with a detectable moiety or therapeutic compound or both bound and conjugated. Such conjugation techniques are well known in the art. For example, conjugation of fluorescent, radioactive or enzymatic moieties can be performed as described in the art (33,43). The detectable moieties contemplated in the present invention can include fluorescent, radioactive and enzymatic markers and the like. Therapeutic drugs contemplated with the present invention can include cytotoxic moieties such as ricin A chain, diphtheria toxin, pseudomonas exotoxin and other chemotherapeutic compounds.

20

It is well understood by one of skill in the art that all of the above discussion regarding antibodies to P/CAF can also be applied with regard to production, characterization and use of antibodies which bind the p300/CBP protein or any of the DNA-binding transcription factors of this invention.

25

30

### Measuring the P/CAF protein in a sample

The present invention also provides a method for determining the presence and thus the amount of P/CAF protein in a biological sample. As used herein, a biological sample includes any tissue or cell which would contain the P/CAF protein. Examples of  
5 cells include tissues taken from surgical biopsies, isolated from a body fluid or prepared in an *in vitro* tissue culture environment.

One example of determining the amount of P/CAF in a biological sample can comprise contacting the biological sample with a polypeptide comprising the amino acid  
10 sequence of SEQ ID NO:3 under conditions whereby a P/CAF/p300 complex can be formed; and determining the amount of the P/CAF/p300 complex, the amount of the complex indicating the amount of P/CAF in the sample. Determination of the amount of P/CAF/p300 complex can be accomplished through techniques standard in the art. For example, the complex may be precipitated out of a solution and detected by the  
15 addition of a detectable moiety conjugated to the p300 protein or by the detection of an antibody which binds p300 or the P/CAF protein, as taught in the Examples herein. Antibodies which bind p300 or the P/CAF protein can be either monoclonal or polyclonal antibodies and can be obtained as described herein. Detection of P/CAF/p300 complexes by the detection of the binding of antibodies reactive with p300  
20 or the P/CAF protein can be accomplished using various immunoassays as are available in the art, as described below.

Alternatively, determination of the amount of P/CAF in a biological sample can comprise contacting the biological sample with a polypeptide comprising the amino acid  
25 sequence of SEQ ID NO:9 under conditions whereby a P/CAF/CBP complex can be formed; and determining the amount of the P/CAF/CBP complex, the amount of the complex indicating the amount of P/CAF in the sample. Determination of the amount of P/CAF/CBP complex can be accomplished through techniques standard in the art. For example, the complex may be precipitated out of a solution and detected by the  
30 addition of a detectable moiety conjugated to the CBP protein or by the detection of an antibody which binds either CBP or the P/CAF protein, as taught in the Examples

herein. Antibodies which bind CBP or the P/CAF protein can be either monoclonal or polyclonal antibodies and can be obtained as described herein. Detection of P/CAF/CBP complexes by the detection of the binding of antibodies reactive with CBP or the P/CAF protein can be accomplished using various immunoassays as are available in the art, as  
5 described below.

Another example of determining the amount of P/CAF in a biological sample comprises contacting the biological sample with an antibody which specifically binds P/CAF under conditions whereby a P/CAF/ antibody complex can be formed and  
10 determining the amount of the P/CAF/antibody complex, the amount of the complex indicating the amount of P/CAF in the sample. Antibodies which bind P/CAF can be either monoclonal or polyclonal antibodies and can be obtained as described herein. Determination of P/CAF/antibody complexes can be accomplished using various immunoassays as are available in the art, as described below.

15

Immunoassays such as immunofluorescence assays, radioimmunoassays (RIA), immunoblotting and enzyme linked immunosorbent assays (ELISA) can be readily adapted for detection and measurement of P/CAF in a biological sample. Both polyclonal and monoclonal antibodies can be used in the assays. Available  
20 immunoassays are well known in the art and are extensively described in the patent scientific literature. See, for example, U.S. Patent Nos. 3,791,932; 3,839,153; 3,850,752; 3,850,578; 3,853,987; 3,867,517; 3,879,262; 3,901,654; 3,935,074; 3,984,533; 3,996,345; 4,034,074; and 4,098,876.

## 25 Screening assays for P/CAF

The present invention also provides a bioassay for screening substances for the ability to inhibit the histone acetyltransferase activity of P/CAF comprising contacting a system, in which histone acetylation by P/CAF can be determined, with the substance under conditions whereby histone acetylation by P/CAF can occur; determining the  
30 amount of histone acetylation by P/CAF in the presence of the substance; and comparing the amount of histone acetylation by P/CAF in the presence of the substance with the

amount of histone acetylation by P/CAF in the absence of the substance, a decreased amount of histone acetylation by P/CAF in the presence of the substance indicating a substance that can inhibit the histone acetyltransferase activity of P/CAF. The acetylation of histones by P/CAF can be determined in a system including, for example, 5 either core histones (histones H2A, H2B, H3 and H4) or the nucleosome core particles (146 base pairs of DNA wrapped around the octamer of core histones) as substrates, the P/CAF protein and radiolabeled acetyl-CoA (e.g., [1-<sup>14</sup>C]acetyl CoA). The presence of acetylated histones can be detected by autoradiography after separation by SDS-PAGE as described herein in the Examples. Thus, the compound to be tested for the ability to 10 inhibit the histone acetyltransferase activity of P/CAF can be added to this system and assayed for inhibiting ability.

The present invention also provides a bioassay for screening substances for the ability to inhibit the transcription modulating activity of P/CAF, comprising contacting a 15 system, in which histone acetylation by P/CAF can be determined, with the substance under conditions whereby histone acetylation by P/CAF can occur; determining the amount of histone acetylation by P/CAF in the presence of the substance; and comparing the amount of histone acetylation by P/CAF in the presence of the substance with the amount of histone acetylation by P/CAF in the absence of the substance, a decreased 20 amount of histone acetylation by P/CAF in the presence of the substance indicating a substance that can inhibit the transcription modulating activity and cell cycle progression suppressing activity of P/CAF. The acetylation of histones by P/CAF can be determined in a system including, for example, either core histones (histones H2A, H2B, H3 and H4) or the nucleosome core particles (146 base pairs of DNA wrapped around the 25 octamer of core histones) as substrates, the P/CAF protein and radiolabeled acetyl-CoA (e.g., [1-<sup>14</sup>C]acetyl CoA). The presence of acetylated histones can be detected by autoradiography after separation by SDS-PAGE as described herein in the Examples. Thus, the compound to be tested for the ability to inhibit the transcription modulating activity of P/CAF by interfering with the histone acetyltransferase activity of P/CAF can 30 be added to this system and assayed for inhibiting ability.

Also provided in the present invention is a bioassay for screening substances for the ability to inhibit the binding of p300 to P/CAF, comprising contacting a system in which the binding of p300 to P/CAF can be determined, with the substance under conditions whereby the binding of p300 and P/CAF can occur; determining the amount of p300 binding to P/CAF in the presence of the substance; and comparing the amount of p300 binding to P/CAF in the presence of the substance with the amount of p300 binding to P/CAF in the absence of the substance, a decreased amount of p300 binding to P/CAF in the presence of the substance indicating a substance that can inhibit the binding of p300 to P/CAF. The binding of p300 to P/CAF can be determined in a system, for example, which can include a cell free reaction mixture comprising a fragment of the p300 protein comprising the amino acid sequence of SEQ ID NO:3 and P/CAF. Alternatively, the system can comprise a cell extract produced from cells producing both p300 and P/CAF. Determination of the binding of p300 to P/CAF can be carried out as taught herein.

15

Additionally provided in the present invention is a bioassay for screening substances for the ability to inhibit the binding of CBP to P/CAF, comprising contacting a system in which the binding of CBP to P/CAF can be determined, with the substance under conditions whereby the binding of CBP to P/CAF can occur; determining the amount of CBP binding to P/CAF in the presence of the substance; and comparing the amount of CBP binding to P/CAF in the presence of the substance with the amount of CBP binding to P/CAF in the absence of the substance, a decreased amount of CBP binding to P/CAF in the presence of the substance indicating a substance that can inhibit the binding of CBP to P/CAF. The binding of CBP to P/CAF can be determined in a system, for example, which can include a cell free reaction mixture comprising a fragment of the CBP protein comprising the amino acid sequence of SEQ ID NO:9 and P/CAF. Alternatively, the system can comprise a cell extract produced from cells producing both CBP and P/CAF. Determination of the binding of CBP to P/CAF can be carried out as taught herein.

30

The present invention further contemplates a bioassay for screening substances for the ability to stimulate the histone acetyltransferase activity of P/CAF comprising contacting a system, in which histone acetylation by P/CAF can be determined, with the substance; determining the amount of histone acetylation by P/CAF in the presence of the substance; and comparing the amount of histone acetylation by P/CAF in the presence of the substance with the amount of histone acetylation by P/CAF in the absence of the substance, an increased amount of histone acetylation by P/CAF in the presence of the substance indicating a substance that can stimulate the histone acetyltransferase activity of P/CAF. The acetylation of histones by P/CAF can be determined in a system including, for example, either core histones (histones H2A, H2B, H3 and H4) or the nucleosome core particles (146 base pairs of DNA wrapped around the octamer of core histones) as substrates, the P/CAF protein and radiolabeled acetyl-CoA (e.g., [1-<sup>14</sup>C]acetyl CoA). The presence of acetylated histones can be detected by autoradiography after separation by SDS-PAGE as described herein in the Examples. Thus, the compound to be tested for the ability to stimulate the histone acetyltransferase activity of P/CAF can be added to this system and assayed for stimulating ability.

The present invention further contemplates a bioassay for screening substances for the ability to stimulate the transcription modulating activity of P/CAF comprising contacting a system, in which histone acetylation by P/CAF can be determined, with the substance; determining the amount of histone acetylation by P/CAF in the presence of the substance; and comparing the amount of histone acetylation by P/CAF in the presence of the substance with the amount of histone acetylation by P/CAF in the absence of the substance, an increased amount of histone acetylation by P/CAF in the presence of the substance indicating a substance that can stimulate the transcription modulating activity of P/CAF. The acetylation of histones by P/CAF can be determined in a system including, for example, either core histones (histones H2A, H2B, H3 and H4) or the nucleosome core particles (146 base pairs of DNA wrapped around the octamer of core histones) as substrates, the P/CAF protein and radiolabeled acetyl-CoA (e.g., [1-<sup>14</sup>C]acetyl CoA). The presence of acetylated histones can be detected by autoradiography after separation by SDS-PAGE as described herein in the Examples.

Thus, the compound to be tested for the ability to stimulate the transcription modulating activity of P/CAF by increasing the histone acetyltransferase activity of P/CAF can be added to this system and assayed for stimulating ability.

5           The present invention further provides a bioassay for screening substances for the ability to stimulate binding of p300 to P/CAF, comprising contacting a system in which the binding of p300 to P/CAF can be determined, with the substance under conditions whereby the binding of p300 to P/CAF can occur; determining the amount of p300 binding to P/CAF in the presence of the substance; and comparing the amount of  
10   p300 binding to P/CAF in the presence of the substance with the amount of p300 binding to P/CAF in the absence of the substance, an increased amount of p300 binding to P/CAF in the presence of the substance indicating a substance that can stimulate the binding of p300 to P/CAF. The binding of p300 to P/CAF can be determined in a system, for example, which can include a cell free reaction mixture comprising a  
15   fragment of the p300 protein comprising the amino acid sequence of SEQ ID NO:3 and P/CAF. Alternatively, the system can comprise a cell extract produced from cells producing both p300 and P/CAF. Determination of the binding of p300 to P/CAF can be carried out as taught herein.

20           Additionally provided in the present invention is a bioassay for screening substances for the ability to stimulate the binding of CBP to P/CAF, comprising contacting a system in which the binding of CBP to P/CAF can be determined, with the substance under conditions whereby the binding of CBP to P/CAF can occur; determining the amount of CBP binding to P/CAF in the presence of the substance; and  
25   comparing the amount of CBP binding to P/CAF in the presence of the substance with the amount of CBP binding to P/CAF in the absence of the substance, an increased amount of CBP binding to P/CAF in the presence of the substance indicating a substance that can stimulate the binding of CBP to P/CAF. The binding of CBP to P/CAF can be determined in a system, for example, which can include a cell free  
30   reaction mixture comprising a fragment of the CBP protein comprising the amino acid sequence of SEQ ID NO:9 and P/CAF. Alternatively, the system can comprise a cell

extract produced from cells producing both CBP and P/CAF. Determination of the binding of CBP to P/CAF can be carried out as taught herein.

### **Transcription modulating activity of P/CAF**

5           The present invention contemplates a method for inhibiting the transcription modulating activity of P/CAF in a subject, comprising administering to the subject a transcription modulating activity inhibiting amount of a substance in a pharmaceutically acceptable carrier. For example, the substance can be identified according to the protocols provided herein as one that can inhibit the transcription modulating activity of  
10 P/CAF by preventing the binding of P/CAF to p300/CBP or by inhibiting the histone acetyltransferase activity of P/CAF as well as by any other inhibitory mechanism as identified by the protocols provided herein. Inhibition of the transcription modulating activity of P/CAF in a subject is desirable, for example, to inhibit HIV TAT-mediated transcription and therefore, the method of the present invention can be used to treat  
15 HIV-infected subjects.

          The substance can be in a pharmaceutically acceptable carrier. By "pharmaceutically acceptable" is meant a material that is not biologically or otherwise undesirable, i.e., the material may be administered to a subject, along with the substance,  
20 without causing any undesirable biological effects or interacting in a deleterious manner with any of the other components of the pharmaceutical composition in which it is contained. The carrier would naturally be selected to minimize any degradation of the active ingredient and to minimize any adverse side effects in the subject.

25           The transcription modulating activity and/or histone acetyltransferase activity of P/CAF can be inhibited in a subject by administering to the subject a substance which binds p300/CBP at the P/CAF binding site or a substance which binds the P/CAF protein at the p300/CBP binding site, the ultimate result being that P/CAF and p300/CBP do not bind with one another and P/CAF cannot exert its transcription  
30 modulating and/or histone acetyltransferase effect. The substance can be a protein, such as an antibody which binds the P/CAF protein binding site at or near the p300/CBP

binding site, thereby preventing its binding or an antibody which binds the p300/CBP protein at or near the P/CAF binding site, thereby preventing its binding. The substance can also bind the histone acetyltransferase site on P/CAF or at the acetylation site on the histone, thereby preventing acetylation by P/CAF.

5

The substance which binds p300/CBP, the P/CAF protein or the histone and has the net effect of inhibiting the transcription modulating effect and or histone acetyltransferase activity of P/CAF in the cell can be delivered to a cell in the subject by mechanisms well known in the art.

10

Alternatively, a nucleic acid encoding a protein which binds either to p300/CBP or the P/CAF protein and has the net effect of inhibiting the transcription modulating effect and/or histone acetyltransferase activity of P/CAF in the cell can be delivered to a cell in the subject by gene transduction mechanisms well known in the art. For example, nucleic acid can be introduced by liposomes as well as via retroviral or adeno-associated viral vectors, as described below.

The substance which inhibits the transcription modulating effect and/or histone acetyltransferase activity of P/CAF can be an antisense RNA or an antisense DNA which binds the RNA or DNA of P/CAF, thereby preventing translation or transcription of the RNA or DNA encoding P/CAF and having the net effect of inhibiting the transcription modulating effect and/or histone acetyltransferase activity of P/CAF by inhibiting P/CAF production. The antisense RNA of the present invention can be generated from the nucleic acid of SEQ ID NO:14 (human) or SEQ ID NO:15 (mouse). Furthermore, the antisense DNA can be a phosphorothioate oligodeoxynucleotide having the nucleotide sequence of SEQ ID NO:16 (human) or of SEQ ID NO:17 (mouse). The mouse antisense RNA can be used to inhibit the activity of mouse P/CAF, having the nucleotide sequence of SEQ ID NO:18 and the amino acid sequence of SEQ ID NO:8. The present invention also contemplates an antisense nucleic acid sequence which can bind the DNA or RNA of any of the transcription factors or other proteins now known or later identified to bind P/CAF, thereby inhibiting expression of the gene products of

30

these proteins and having the net effect of inhibiting the transcription modulating effect and/or histone acetyltransferase activity of P/CAF.

The antisense nucleic acid can comprise a typical nucleic acid, but the antisense  
5 nucleic acid can also be a modified nucleic acid or a derivative of a nucleic acid such as a phosphorothioate analogue of a nucleic acid. The composition can comprise, for example, an antisense RNA that specifically binds an RNA encoded by the gene encoding the serum protein. Antisense RNAs can be synthesized and used by standard methods (62).

10

Antisense RNA can inhibit gene expression by forming an RNA/RNA duplex between the antisense RNA and the RNA transcribed from the target gene. The precise mechanism by which this duplex formation decreases the production of the protein encoded by the endogenous gene probably involves binding of complementary regions  
15 of the normal sense mRNA and the antisense RNA strand with duplex formation in a manner that blocks RNA processing and translation. Alternative mechanisms include the formation of a triplex between the antisense RNA and duplex DNA or the formation of an DNA-RNA duplex with subsequent degradation of DNA-RNA hybrids by RNase H. Furthermore, an antigene effect can result from certain DNA-based oligonucleotides  
20 via triple-helix formation between the oligomer and double-stranded DNA which results in the repression of gene transcription. Regardless of the specific molecular mechanism, the present invention results in inhibition of expression of the P/CAF gene by the introduced and replicated DNA resulting in inhibition of the transcription modulating and/or histone acetyltransferase activity of P/CAF, by a reduction in the expression of  
25 the nucleic acid to which the antisense nucleic acid is hybridized, and therefore a reduction of the gene product from the targeted gene.

The antisense nucleic acid may be obtained by any number of techniques known to one skilled in the art. One method of constructing an antisense nucleic acid is to  
30 synthesize a recombinant antisense DNA molecule. For example, oligonucleotide synthesis procedures are routine in the art and oligonucleotides coding for a particular

protein or regulatory region are readily obtainable through automated DNA synthesis. A nucleic acid for one strand of a double-stranded molecule can be synthesized and hybridized to its complementary strand. One can design these oligonucleotides such that the resulting double-stranded molecule has either internal restriction sites or appropriate 5' or 3' overhangs at the termini for cloning into an appropriate vector. Double-stranded molecules coding for relatively large proteins or regulatory regions can be synthesized by first constructing several different double-stranded molecules that code for particular regions of the protein or regulatory region, followed by ligating these DNA molecules together. Once the appropriate DNA molecule is synthesized, this DNA can be cloned downstream of a promoter in an antisense orientation. Techniques such as this are routine in the art and are well documented.

An example of another method of obtaining an antisense nucleic acid is to isolate that nucleic acid from the organism in which it is found and clone it in an antisense orientation. For example, a DNA or cDNA library can be constructed and screened for the presence of the nucleic acid of interest. Methods of constructing and screening such libraries are well known in the art and kits for performing the construction and screening steps are commercially available (for example, Stratagene Cloning Systems, La Jolla, CA). Once isolated, the nucleic acid can be directly cloned into an appropriate vector in an antisense orientation, or if necessary, be modified to facilitate the subsequent cloning steps. Such modification steps are routine, an example of which is the addition of oligonucleotide linkers which contain restriction sites to the termini of the nucleic acid. General methods are set forth in Sambrook *et al.* (39).

The DNA that is introduced into the cell is in an expression orientation that is antisense to a corresponding endogenous DNA or RNA of the cells. For example, where an endogenous DNA comprises a gene which encodes for a particular protein, the introduced DNA is in an expression orientation opposite the expression of the endogenous DNA; that is the DNA operatively linked to a promoter is in an antisense expression orientation relative to the corresponding endogenous gene. The introduced DNA may be homologous to the entire transcribed gene or homologous to only part of

the transcribed gene. Alternatively, the sequence of the introduced DNA may be divergent to that of the endogenous DNA but only divergent to the extent that hybridization of the nucleic acids occurs, thereby preventing transcription. One skilled in the art can determine the maximum extent of this divergence by routine screening of antisense DNAs corresponding to an endogenous DNA of the cell. In this manner, one skilled in the art can readily determine which fragments, or alternatively the extent of homology of the fragments or the entire gene that is necessary to inhibit gene expression.

10           The antisense nucleic acids of the present invention can be made according to protocols standard in the art, as well as described in the Examples provided herein. The antisense nucleic acids can be administered to a subject according to the gene transduction protocols standard in the art, as described below.

15           The present invention also contemplates a method for stimulating the transcription modulating activity and/or histone acetyltransferase activity of P/CAF in a subject comprising administering to the subject a substance, in a pharmaceutically acceptable carrier, determined according to the methods taught herein, to have a stimulatory affect on the transcription modulating and/or histone acetyltransferase activity of P/CAF. The substance can be one which has been identified, according to the protocols provided herein, to stimulate histone acetyltransferase activity in P/CAF or promote binding of P/CAF to p300/CBP. The stimulation of the transcription modulation activity and/or histone acetyltransferase activity of P/CAF in a subject is desirable, for example, to activate tumor suppressor p53 (which promotes apoptosis) or to activate the muscle differentiation factor, MyoD. Thus, the method of the present invention can be employed to treat cancer and to promote muscle differentiation in conditions where muscle differentiation is desired. The substance can be delivered to a cell in the subject by mechanisms well known in the art.

30           Further contemplated in the present invention is a method for promoting binding of P/CAF to p300/CBP in a subject, comprising administering to the subject a substance

identified by the methods provided herein to promote binding of P/CAF to either p300 or CBP.

Additionally, a nucleic acid encoding a protein which stimulates the transcription  
5 modulating activity and/or histone acetyltransferase activity of P/CAF can be delivered to a cell in the subject by gene transduction mechanisms, as described below.

Also provided in the present invention is a method of inhibiting the cell cycle  
progression inducing effect of an oncoprotein which binds p300/CBP in a subject  
10 comprising transducing the cells of the subject with a vector comprising a nucleic acid encoding the P/CAF protein; inducing expression of the nucleic acid in the cell to produce the P/CAF in an amount which will allow the P/CAF gene product to replace the oncoprotein bound to p300/CBP, whereby the replacement of the oncoprotein bound to p300/CBP by the P/CAF gene product inhibits the cell cycle progression  
15 inducing effect of the oncoprotein. The oncoprotein which binds p300/CBP in the cell can be the adenovirus E1A oncoprotein.

A method for providing a functional P/CAF protein to a subject in need of the functional P/CAF protein is also provided, comprising transducing the cells of the  
20 subject with a vector comprising a nucleic acid encoding the P/CAF protein and inducing expression of the nucleic acid to produce the functional P/CAF protein in the cell, thereby providing the functional P/CAF protein to the subject. The transduction of the vector nucleic acid into the subject's cells can be carried out according to standard gene therapy protocols well known in the art (see, for example, U.S. Patent No.  
25 5,339,346).

#### **Screening assays for p300/CBP**

The present invention also provides a bioassay for screening substances for the ability to inhibit the histone acetyltransferase activity of p300/CBP comprising  
30 contacting a system, in which histone acetylation by p300/CBP can be determined, with the substance under conditions whereby histone acetylation by p300/CBP can occur;

determining the amount of histone acetylation by p300/CBP in the presence of the substance; and comparing the amount of histone acetylation by p300/CBP in the presence of the substance with the amount of histone acetylation by p300/CBP in the absence of the substance, a decreased amount of histone acetylation by p300/CBP in the presence of the substance indicating a substance that can inhibit the histone acetyltransferase activity of p300/CBP. The acetylation of histones by p300/CBP can be determined in a system including, for example, either core histones (histones H2A, H2B, H3 and H4) or the nucleosome core particles (146 base pairs of DNA wrapped around the octamer of core histones) as substrates, the P300/CBP protein and radiolabeled acetyl-CoA (e.g., [1-<sup>14</sup>C]acetyl CoA). The presence of acetylated histones can be detected by autoradiography after separation by SDS-PAGE as described herein in the Examples. Thus, the compound to be tested for the ability to inhibit the histone acetyltransferase activity of p300/CBP can be added to this system and assayed for acetyltransferase inhibiting ability.

15

Also provided in the present invention is a bioassay for screening substances for the ability to inhibit the binding of a transcriptional factor to p300/CBP, comprising contacting a system in which the binding of a transcriptional factor to p300/CBP can be determined, with the substance under conditions whereby the binding of the transcriptional factor and p300/CBP can occur; determining the amount of transcriptional factor binding to p300/CBP in the presence of the substance; and comparing the amount of transcriptional factor binding to p300/CBP in the presence of the substance with the amount of transcriptional factor binding to p300/CBP in the absence of the substance, a decreased amount of transcriptional factor binding to p300/CBP in the presence of the substance indicating a substance that can inhibit the binding of a transcriptional factor to p300/CBP. The binding of a transcriptional factor to p300/CBP can be determined in a system, for example, which can include a cell free reaction mixture comprising a transcriptional factor which binds p300/CBP and p300/CBP. Alternatively, the system can comprise a cell extract produced from cells producing both a transcriptional factor which binds p300/CBP and p300/CBP. The transcriptional factor which binds p300/CBP can be selected from, but is not limited to

30

the group consisting of nuclear hormone receptors, CREB, c-Jun/v-Jun, c-Myb/v-Myb, YYI, Sap-1a, c-Fos, MyoD and SRC-1, as well as any other transcriptional factor now known or later identified to bind p300/CBP. The screening assay of the present invention can also be used to identify substances which inhibit the binding of p300/CBP to other components to which it is known to bind, for example, P/CAF, pp90<sub>RSK</sub>, TFIIB, E1A, SV40 large T antigen, as well as any other substances now known or later identified to bind p300/CBP. Determination of the binding of a transcriptional factor or other substance to p300/CBP can be carried out as taught in the Examples herein as well as by protocols described in the literature.

10

The present invention further contemplates a bioassay for screening substances for the ability to stimulate the histone acetyltransferase activity of p300/CBP comprising contacting a system, in which histone acetylation by p300/CBP can be determined, with the substance; determining the amount of histone acetylation by p300/CBP in the presence of the substance; and comparing the amount of histone acetylation by p300/CBP in the presence of the substance with the amount of histone acetylation by p300/CBP in the absence of the substance, an increased amount of histone acetylation by p300/CBP in the presence of the substance indicating a substance that can stimulate the histone acetyltransferase activity of p300/CBP. The acetylation of histones by p300/CBP can be determined in a system including, for example, either core histones (histones H2A, H2B, H3 and H4) or the nucleosome core particles (146 base pairs of DNA wrapped around the octamer of core histones) as substrates, the p300/CBP protein and radiolabeled acetyl-CoA (e.g., [1-<sup>14</sup>C]acetyl CoA). The presence of acetylated histones can be detected by autoradiography after separation by SDS-PAGE as described herein in the Examples. Thus, the compound to be tested for the ability to stimulate the histone acetyltransferase activity of p300/CBP can be added to this system and assayed for stimulating ability.

25

The present invention further provides a bioassay for screening substances for the ability to stimulate binding of a component, which binds p300/CBP, to p300/CBP, comprising contacting a system in which the binding of the component to p300/CBP can

30

be determined, with the substance under conditions whereby the binding of the component to p300/CBP can occur; determining the amount of component binding to p300/CBP in the presence of the substance; and comparing the amount of component binding to p300/CBP in the presence of the substance with the amount of component binding to p300/CBP in the absence of the substance, an increased amount of component binding to p300/CBP in the presence of the substance indicating a substance that can stimulate the binding of the component to p300/CBP. The binding of the component to p300/CBP can be determined in a system, for example, which can include a cell free reaction mixture comprising the component and p300/CBP. Alternatively, the system can comprise a cell extract produced from cells producing both the component and p300/CBP. The component which binds p300/CBP can be any of the transcriptional factors or other proteins which are known or are identified in the future to bind p300/CBP, as set forth above. Determination of the binding of the component to p300/CBP can be carried out as taught in the Examples provided herein and according to protocols available in the literature.

#### **Histone acetyltransferase activity of p300/CBP**

A method for inhibiting the histone acetyltransferase activity of p300/CBP in a subject is provided in the present invention, comprising administering to the subject a histone acetyltransferase activity inhibiting amount of a substance in a pharmaceutically acceptable carrier. The mechanism of the inhibitory action of the substance can be the inhibition of the binding of a DNA-binding transcription factor, such as, for example, a nuclear hormone receptor, CREB, c-Jun/v-Jun, c-Myb/v-Myb, YY1, Sap-1a, c-Fos, MyoD or SRC-1, to p300/CBP.

25

The histone acetyltransferase activity of p300/CBP can be inhibited in a subject by administering to the subject a substance which binds p300/CBP at the transcription factor binding site or a substance which binds the transcription factor protein at the p300/CBP binding site, the ultimate result being that the transcription factor and p300/CBP do not bind with one another and p300/CBP cannot acetylate histones.

30

The substance which binds either to the transcription factor or the p300/CBP protein and has the net effect of inhibiting the histone acetyltransferase activity of p300/CBP in the cell can be identified according to the screening methods provided herein and delivered to a cell in the subject by mechanisms well known in the art. The substance can be a protein, such as an antibody which binds the p300/CBP protein binding site at or near the DNA-binding transcription factor binding site, thereby preventing its binding or an antibody which binds the DNA-binding transcription factor at or near the p300/CBP binding site, thereby preventing its binding. The substance can also bind the histone acetyltransferase site on p300/CBP (aa 1195-1673 on p300 or aa 1174-1850 on CBP) or at the acetylation site on the histone, thereby preventing acetylation by p300/CBP.

Additionally, the substance can be a nucleic acid which can be expressed in the cell to produce a protein which inhibits the histone acetyltransferase activity of p300/CBP. For example, a nucleic acid encoding a protein which binds either to a transcription factor or the p300/CBP protein and has the net effect of inhibiting the histone acetyltransferase activity of p300/CBP in the cell can be delivered to a cell in the subject by gene transduction mechanisms well known in the art. For example, nucleic acid can be introduced by liposomes as well as via retroviral or adeno-associated viral vectors, as described below.

The substance which inhibits the histone acetyltransferase activity of p300/CBP can be an antisense RNA or an antisense DNA which binds the RNA or DNA of p300/CBP thereby preventing translation or transcription of the RNA or DNA encoding p300/CBP and having the net effect of inhibiting the histone acetyltransferase activity of P/CAF by inhibiting p300/CBP production. The antisense RNA or DNA of the present invention can be produced and introduced into cells according to the same methods as set forth above for P/CAF antisense nucleic acids.

The present invention also contemplates a method for stimulating the histone acetyltransferase activity of p300/CBP in a subject comprising administering to the

subject a histone acetyltransferase activity stimulating amount of a substance, in a pharmaceutically acceptable carrier, determined according to the methods taught herein, to have a stimulatory affect on the histone acetyltransferase activity of p300/CBP. The substance can exert a stimulatory effect by promoting the binding of a DNA-binding transcription factor of the present invention to p300/CBP. The substance can be delivered to a cell in the subject by mechanisms well known in the art. A nucleic acid encoding a protein which stimulates the transcription modulating activity of p300/CBP can be delivered to a cell in the subject by gene transduction mechanisms, as described below.

10

### **Gene transduction**

In the methods described above which include gene transduction into cells (i.e., addition of exogenous DNA into cells), the nucleic acids of the present invention can be in a vector for delivering the nucleic acids to the site for expression of the P/CAF protein. The vector can be one of the commercially available preparations, such as the pGM plasmid (Promega). Vector delivery can be by liposome, using commercially available liposome preparations or newly developed liposomes having the features of the present liposomes. Additionally, vector delivery can be via a viral system, including, but not limited to, retroviral, adenoviral and adeno-associated viral systems. Other delivery methods can be adopted and routinely tested according to the methods taught herein.

The modes of administration of the liposome will vary predictably according to the disease being treated and the tissue being targeted. For example, for treating cancer in either the lung or the liver, which are both sinks for liposomes, intravenous delivery is reasonable. For other localized cancers, as well as precancerous conditions, catheterization of an artery upstream from the target organ is a preferred mode of delivery, because it avoids significant clearance of the liposome by the lung and liver. For cancerous lesions at a number of other sites (e.g., skin cancer, localized dysplasias), topical delivery is expected to be effective and may be preferred, because of its convenience.

30

Leukemias and other disorders involving dysregulated proliferation of certain isolatable cell populations may be more readily treated by *ex vivo* administration of the nucleic acid.

5           The liposomes may be administered topically, parenterally (e.g., intravenously), by intramuscular injection, by intraperitoneal injection, transdermally, extracorporeally or the like, although intravenous or topical administration is typically preferred. The exact amount of the liposomes required will vary from subject to subject, depending on the species, age, weight and general condition of the subject, the severity of the disease  
10   being treated, the particular compound used, its mode of administration and the like. Thus, it is not possible to specify an exact amount. However, an appropriate amount may be determined by one of ordinary skill in the art using only routine experimentation given the teachings herein.

15           Parenteral administration, if used, is generally characterized by injection. Injectables can be prepared in conventional forms, either as liquid solutions or suspensions, solid forms suitable for solution or suspension in liquid prior to injection, or as emulsions. A more recently revised approach for parenteral administration involves use of a slow release or sustained release system such that a constant level of dosage is  
20   maintained. See, e.g., U.S. Patent No. 3,610,795, which is incorporated by reference herein.

Topical administration can be by creams, gels, suppositories and the like. *Ex vivo* (extracorporeal) delivery can be as typically used in other contexts.

25

The present invention is more particularly described in the following examples which are intended as illustrative only since numerous modifications and variations therein will be apparent to those skilled in the art.

30

## EXAMPLES

### I. P/CAF studies.

#### 5 Cloning and characterization of P/CAF protein.

In human cells, CBP binds to c-Jun in a phosphorylation-dependent manner in association with stimulation of transcription (9). In yeast, GCN4 is believed to be a c-Jun counterpart on the basis of similarities in DNA recognition (15) as well as the participation of both proteins in UV signaling pathways (16). Yeast genetic screening  
10 has led to the isolation of various cofactors for GCN4, including GCN5 (yGCN5), ADA2 (yADA2) and ADA3 (yADA3) (17-19). These factors are considered to function as a complex (or in a common pathway) based on genetic and protein-protein interaction studies (18-22). Finally, p300/CBP and yADA2 exhibit significant sequence similarity within a 50 amino acid region including a Zn<sup>2+</sup> finger motif (3). Human  
15 counterparts to yGCN5, yADA2, or yADA3 that interact with p300/CBP to mediate transcriptional activation by c-Jun were searched for in various nucleotide sequence databases.

Comparison of the yGCN5 protein sequence with various databases (23)  
20 revealed significant similarities with the two randomly sequenced human cDNAs, ETS05039 (24) ( $P=4.0 \times 10^{-15}$ ) and NIB2000-5R ( $P=6.5 \times 10^{-9}$ ). Given that these cDNAs were truncated, human fetal liver and fetal brain cDNA libraries (Clontech) were screened with ETS05039 and NIB2000-5R, respectively and complete clones were isolated from the human fetal liver cDNA library. The complete sequences revealed that  
25 the ETS05039- and NIB2000-5R-derived clones are encoded by distinct genes but are highly related within the protein coding regions (68% identity at the DNA level; 75% identity and 86% similarity at the protein level). The former encodes an N-terminal region with no sequence similarity to any proteins in the databases besides the yGCN5-related C-terminal region, whereas the latter encodes only the yGCN5-related region.  
30 Given that p300/CBP-binding activity was observed in the former polypeptide as shown below, it was designated p300/CBP-associated factor (P/CAF), having the amino acid

sequence of SEQ ID NO:1 and the nucleotide sequence of SEQ ID NO:10 and the latter was named human GCN5 (hGCN5), having the amino acid sequence of SEQ ID NO:5 and the nucleotide sequence of SEQ ID NO:11.

5        Additionally, an RNA blot (Clontech) was hybridized with a random-primed probe made from the cDNA encoding P/CAF. RNA blotting indicated that transcripts detected by the P/CAF and hGCN5 cDNAs are ubiquitously expressed, but the former is most abundant in heart and skeletal muscle, whereas the latter is most abundant in pancreas and skeletal muscle.

10

#### **P/CAF-p300/CBP interaction *in vitro***

      The P/CAF binding site was presumed to reside in the C terminal one third of CBP (residues 1,678-2,442) because it was observed that this region, when fused to a DNA binding domain, activates transcription (4) in a manner repressed by coexpression  
15 of 12S E1A. This region was divided into 6 overlapping fragments and each was expressed in *E. coli* as a glutathione-S-transferase (GST) fusion protein. GST-CBP fusions were incubated with recombinant P/CAF protein and, subsequently, purified using glutathione-Sepharose. Co-purified P/CAF was detected by immunoblotting analysis.

20

      To construct GST-fusions, various regions of CBP and p300 were amplified by PCR. A series of deletions of the CBP segment B was created by site-directed *in vitro* mutagenesis (30). These fragments were subcloned into pGEX-2T (Pharmacia). GST-fusions were expressed in *E. coli* and extracted with buffer B [20 mM Tris-HCl (pH  
25 8.0), 5 mM MgCl<sub>2</sub>, 10% glycerol, 1 mM AEBSF, 0.1% NP40, 10 µg/ml of aprotinin, 10 µg/ml of leupeptin, 1 µg/ml of pepstatin A, 1 mM DTT] containing 0.1 M KCl for these experiments. GST-CBP-segment B was purified by glutathione-Sepharose and phenyl-Sepharose chromatographic steps. P/CAF, hGCN5, and E1A were expressed as FLAG-fusions in Sf9 cells via baculovirus vectors and affinity-purified with M2-agarose (ref.  
30 30; Kodak-IBI). For interaction, a crude *E. coli* extract containing 20 pmol of GST-fusion was incubated with 40-60 pmol of P/CAF or E1A in a total volume of 50 µl of

buffer B with 0.1 M KCl on ice for 10 min. Samples were further incubated with 10  $\mu$ l (packed volume) of glutathione-Sepharose at 4°C for 30 min, washed four times with 200  $\mu$ l of buffer B containing 0.1 M KCl, and eluted with 20  $\mu$ l of buffer E [50 mM Tris-HCl (pH 8.0), 0.2 M KCl, 20 mM glutathione] for 60 min. Interacting proteins  
5 were detected by anti-FLAG immunoblotting or silver staining.

For p300 interactions, the segment spanning residues 1763-1966 (segment B') of p300, which is analogous to the CBP segment-B, was used. Twenty percent of the P/CAF and hGCN5 inputs and 100% of the E1A input were also analyzed. In the GST  
10 precipitation assays, almost identical amounts of the GST fusions were recovered in all samples. Interaction between P/CAF and CBP (segment B) was determined in the absence and in the presence of E1A. Control reactions with GST-CBP alone and without GST-CBP were also performed. Input proteins were analyzed.

15 Two CBP segments, A and B, interacted specifically with P/CAF. The stronger interaction was observed in the latter segment, which does not include the yADA2-like Zn<sup>2+</sup> finger. Given that the CBP segment-B is well conserved in p300 (66% identity, 75% similarity), the binding of P/CAF to p300 *in vitro* was also analyzed. For this experiment, the p300 segment spanning residues 1763-1966, termed segment B', which  
20 is analogous to the CBP segment-B, was used. Like CBP, p300 interacted specifically with P/CAF. These studies demonstrated that P/CAF binds specifically to both p300 and CBP *in vitro*. In contrast to P/CAF, hGCN5 did not bind to CBP or p300.

These studies also demonstrated that the Zn<sup>2+</sup> finger region of p300/CBP, which  
25 shares sequence similarity with yADA2, is not essential for the interaction with P/CAF. Cloning of a human structural homolog of yADA2, termed hADA2 (25) has revealed that, unlike the sequence similarity between p300/CBP and yADA2, which is restricted to a 50 amino acid region, hADA2 shares extensive similarity (30% identity, 52% similarity) to yADA2 over the entire protein sequence. Moreover, a computer search of  
30 the complete genomic sequence of *Saccharomyces cerevisiae* revealed that yeast does

not have counterparts of p300/CBP or P/CAF. Thus, the p300/CBP-P/CAF pathway may have been acquired during metazoan evolution.

## 5    **Action of E1A *in vitro***

Previous reports indicated that E1A binds to both the p300 segment spanning residues 1767-1816 and the CBP segment spanning residues 1805-1854 (7). These interactions were reconfirmed in the present system; thus, both p300 and CBP segments covering the previously identified regions interacted with E1A.

10

For further mapping, a series of deletions was introduced within the CBP segment-B and tested for interactions with P/CAF and E1A. Deletions of residues 1801-1825 or 1824-1851 markedly reduced interactions with both P/CAF and E1A, whereas deletion of residues 1850-1878 did not affect these interactions. Furthermore, deletion of residues 1801-1851 completely abolished interactions with both P/CAF and E1A. These data indicate that residues 1801-1851 of CBP are critical for interaction with both P/CAF and E1A. Taken together with the evidence that CBP segment A (aa residues 1,678-1,880) also binds to these factors, the above findings demonstrate that P/CAF and E1A bind to the same or very closely spaced sites on CBP.

20

Evidence that both P/CAF and E1A recognize the same p300/CBP segments raises the possibility of direct competition between P/CAF and E1A for binding to p300/CBP. To test this possibility, a competition experiment was performed with the use of affinity purified recombinant proteins. The interaction of P/CAF with the CBP-segment B was progressively inhibited by the addition of increasing amounts of E1A. In contrast, no inhibition was caused by an E1A mutant which does not bind to p300/CBP (E1A $\Delta$ N). Similar results were obtained with the p300-segment B', leading to the conclusion that P/CAF and E1A compete for the same binding sites in p300/CBP.

30

**P/CAF-p300/CBP interaction *in vivo***

The *in vivo* interaction between P/CAF and p300/CBP was established by co-immunoprecipitation from a human osteosarcoma cell extract. Proteins in this extract were immunoprecipitated with rabbit anti-P/CAF, rabbit anti-CBP and anti-p300  
5 antibodies. For controls, cell extract was precipitated with rabbit control IgG or mouse anti-HA monoclonal antibody. The precipitates were analyzed by immunoblotting with anti-P/CAF, anti-CBP and anti-p300 antibodies.

Osteosarcoma cells were transfected with either control vector or E1A- or  
10 E1AΔN-expression vectors. Extract from the transfected subpopulation was immunoprecipitated with anti-P/CAF or control IgG. The precipitates were analyzed by immunoblotting with anti-p300 and anti-P/CAF antibodies.

Rabbit anti-P/CAF antibody was raised to the P/CAF segment spanning residues  
15 125-397 and purified by immunoaffinity chromatography (33). A mixture of monoclonal antibodies raised to the human p300 segment spanning residues 1572-2371 (5) and rabbit polyclonal antibodies raised to the mouse CBP segment spanning residues 2-23 (for immunoprecipitation) and 1736-2179 (immunoblotting) were purchased from Upstate Biotechnology. Approximately  $2 \times 10^7$  human osteosarcoma U-2 OS cells  
20 (ATCC accession number HTB 96) were extracted with 10 ml of lysis buffer [25 mM HEPES-KOH (pH 7.2), 150 mM potassium acetate, 2 mM EDTA, 1 mM DTT, 1 mM AEBSF, 10 µg/ml of aprotinin, 10 µg/ml of leupeptin, 1 µg/ml of pepstatin A, 20 mM sodium fluoride, 0.1% NP40]. Two to 10 ml of extract were incubated with 2 µg of the respective antibody for four hours at 4°C. Fifty µl (packed volume) of protein-A  
25 Trisacryl (Pierce) were added and incubation was continued for two hours. The matrix was washed four times with 1 ml of the lysis buffer, then boiled in 2x SDS sample buffer. Human osteosarcoma U-2 OS cells were transfected with 20 µg of the indicated plasmid and 1 µg of sorting plasmid (pCMV-IL2R)(31). The transfected subpopulation was purified by magnetic affinity cell sorting(32). Extract from approximately  $2 \times 10^5$   
30 sorted cells was immunoprecipitated as described.

Anti-P/CAF antibody specifically detected a 95 kDa protein, which is very close to the calculated value for the full-length P/CAF, in the immunoprecipitates. Anti-P/CAF antibody co-immunoprecipitated both CBP and p300. Similarly, anti-CBP antibody also co-immunoprecipitated P/CAF. However, anti-p300 antibody did not co-immunoprecipitate P/CAF. This is most likely due to steric interference since the anti-p300 antibody was raised to the p300 segment spanning residues 1572-2371 which includes the P/CAF binding region. These data demonstrate that P/CAF forms complexes with both p300 and CBP *in vivo*.

#### 10 Action of E1A *in vivo*

The *in vitro* experiments described herein indicate that P/CAF and E1A compete for the binding sites in p300/CBP. Thus, a study was conducted to determine whether E1A targets the endogenous interaction between P/CAF and p300. An E1A-expression vector was transiently transfected into human osteosarcoma cells and the transfected subpopulation was purified by cell sorting. Then, the interaction between P/CAF and p300 in transfected cells was examined by co-immunoprecipitation with anti-P/CAF antibody. The endogenous interaction of P/CAF with p300 was drastically inhibited by expression of E1A. On the other hand, no inhibition was observed by the E1A mutant lacking the p300 binding domain (E1A $\Delta$ N), indicating that E1A disrupts the P/CAF-p300 complex *in vivo* through an interaction with p300.

#### Cell cycle regulation by P/CAF

Given that binding of P/CAF to p300/CBP is inhibited by E1A, experiments were performed to evaluate whether P/CAF, by binding to and forming a functional complex with p300, is involved in the regulation of entry into S phase. This possibility was addressed by examining whether transient expression of P/CAF would affect the rate of G1/S transit in HeLa cells. P/CAF negatively affected the distribution of cells between G1 and S phases in this assay.

HeLa cells were transfected by electroporation with 7  $\mu$ g of P/CAF-expression plasmid and/or 3  $\mu$ g of the full-length or the N-terminally deleted ( $\Delta$ 2-36) E1A 12S-

expression plasmid as indicated. These plasmids were constructed by subcloning FLAG-P/CAF and E1A cDNAs into pCX (34) and pcDNAI (Invitrogen), respectively. All samples, in addition, contained 1  $\mu$ g of sorting plasmid (pCMV-IL2R) (31) and carrier plasmid (pCX) to normalize the total amount of DNA to 11  $\mu$ g. After  
5 transfection, cells were incubated in Dulbecco's modified Eagle's medium with 10% fetal bovine calf serum for 12 h, and subsequently labeled in medium containing 10  $\mu$ M bromo-deoxyuridine (BrdU) for 30 min. Subsequently, the transfected subpopulation was purified by magnetic affinity cell sorting and nuclei were analyzed by dual parameter flow cytometry as described (32).

10

The fraction of cells accumulating in S phase in control cultures was 23%, compared to 15% in P/CAF-transfected cells. This effect was reproducible in multiple independent experiments. In parallel experiments to verify the utility of this experimental protocol, plasmids encoding E2F-1, simian virus 40 small t, cyclin A or  
15 cyclin E increased the accumulation of cells in S phase, whereas plasmids encoding the cyclin-dependent kinase inhibitors p21 or p27 reduced the number of S phase cells.

On the basis of evidence that E1A and P/CAF compete for binding sites on p300, it seemed possible that cotransfection of P/CAF with E1A would oppose the  
20 mitogenic effect caused by E1A. As shown by the data herein, this is indeed the case. E1A alone has mitogenic activity in this experimental setting, while the E1A mutant lacking the p300 binding domain (E1A $\Delta$ N) has very weak activity. Comparable expression levels between wild type and mutant E1A in the transfected cells were revealed by immunoblotting analysis with anti-E1A. Intriguingly, when P/CAF was  
25 cotransfected with E1A, the mitogenic activity of E1A was significantly counteracted by P/CAF. These results show that P/CAF and E1A mediate antagonistic effects on cell cycle progression.

In the course of assessing P/CAF activity, it was also revealed that p300 is able  
30 to inhibit cell cycle progression under the same assay conditions. These findings suggest

that P/CAF and p300, perhaps by forming a complex, act in concert to suppress cell cycle progression.

#### **Histone acetyltransferase activity in P/CAF**

5           Acetylation of the N-terminal histone tails has been considered to play a crucial role in accessibility of transcription factors to nucleosomal templates (26-27). Recently, yGCN5 has been identified as a histone acetyltransferase (28). On the basis of this information, intrinsic histone acetyltransferase activity in P/CAF and hGCN5 was examined. As substrates, the core histones (histones H2A, H2B, H3 and H4) and the  
10   nucleosome core particles (146 base pairs of DNA wrapped around the octamer of core histones) were used.

          Activity of hGCN5 and P/CAF that acetylates free histones or histones in the nucleosome core particle (35) was measured as described (36). Each reaction contained  
15   0.3 pmol of affinity purified FLAG-hGCN5 or FLAG-P/CAF, 4 pmol of the histone octamer or the nucleosome core particle and 10 pmol of [1-<sup>14</sup>C]acetyl-CoA. The histone octamer dissociated into dimers or tetramers under assay conditions. Acetylated histones were detected by autoradiography after separation by SDS-PAGE.

20           P/CAF and hGCN5 acetylated the core histones with almost the same efficiency. Both factors acetylated histones H3 and H4, but preferentially H3. In contrast, very weak or no acetylation by hGCN5 was detected in the nucleosome core particles. Remarkably, significant acetylation by P/CAF was observed in a nucleosomal context. Although all core histones are acetylated in the nucleus, P/CAF and hGCN5 did not  
25   acetylate histones H2A and H2B *in vitro*.

          Direct function of P/CAF is likely to involve its intrinsic histone acetyltransferase activity. Although exact molecular mechanisms by which acetylation of core histones contribute to transcription remains undefined, acetylation of the histones is considered to  
30   play an important role in transcriptional regulation (26-27). The positively charged N-terminal tails of core histones are believed to affect nucleosome structure by interacting

with DNA at or near the nucleosome-spacer junction. Acetylation of the histone tails presumably destabilizes the nucleosome and facilitates access by regulatory factors. Likewise, there is a general correlation between the level of acetylation and transcriptional activity of nucleosomal domains. The findings of the present invention provide insights into the mechanisms of targeted histone acetylation.

Cellular factor p300/CBP binds to various sequence-specific factors that are involved in cell growth and/or differentiation, including CREB (3,4), c-Jun (9), Fos (11), c-Myb (12) and nuclear receptors (13). P/CAF could stimulate the activation function of these factors via promoter-specific histone acetylation. The present invention demonstrates that E1A appears to perturb normal cellular regulation by disrupting the connection between p300/CBP and its associated histone acetyltransferase.

## **II. p300/CBP studies.**

15

### **Purification of E1A associated histone acetyltransferase.**

FLAG-epitope tagged E1A (or  $\Delta$ E1A) was expressed in Sf9 cells (ATCC accession number CRL 1711) by infecting recombinant baculovirus (43). All purification steps were carried out at 4°C. Extract was prepared from infected cells by one cycle of freeze and thaw in buffer B (20 mM Tris-HCl, pH 8.0; 5 mM MgCl<sub>2</sub>; 10% glycerol; 1 mM PMSF; 10 mM  $\beta$ -mercaptoethanol; 0.1% Tween 20) containing 0.1 M KCl and the complete protease inhibitor cocktail (Boehringer Mannheim). To prepare E1A-immobilized beads, the extract was incubated with M2 anti-FLAG antibody agarose (Kodak-IBI) for four hours with rotating and subsequently washed with the same buffer three times. The resulting beads were incubated with HeLa (ATCC accession number CCL 2) nuclear extract for four to eight hours and thereafter washed with the same buffer six times. Finally, FLAG-E1A was eluted from the beads along with associated polypeptides by incubating with the same buffer containing 0.1 mg/ml FLAG peptide.

30

For further purification, eluted polypeptides were dialyzed in 0.05 M KCl-buffer B and subsequently loaded onto a SMART Mono Q column (Pharmacia) equilibrated with the same 0.05 M KCl-buffer B. After washing, the column was developed with a linear gradient of 0.05-1.0 M KCl in buffer B. Mono Q fractions were concentrated with a MICROCON spin-filter (Amicon) and consequently loaded onto a SMART Superdex 200 column (Pharmacia) equilibrated with 0.1 M KCl-buffer B.

### **Histone acetyltransferase assays**

Filter binding assays were performed as described (80) with minor modifications.

10 Samples were incubated at 30°C for 10-60 minutes in 30 ml of assay buffer containing 50 mM Tris-HCl, pH 8.0; 10% glycerol; 1 mM DTT; 1 mM PMSF; 10 mM sodium butyrate; 6 pmol of [<sup>3</sup>H]acetyl CoA (4.3 mCi/mmol, Amersham Life Science Inc.); and 33 mg/ml of calf thymus histones (Sigma Chemical Co.). In experiments where synthetic peptides were substituted for core histones, 50 pmol of each peptide were used. After

15 incubation, the reaction mixture was spotted onto Whatman P-81 phosphocellulose filter paper and washed for 30 minutes with 0.2 M sodium carbonate buffer pH 9.2 at room temperature with 2-3 changes of the buffer. The dried filters were counted in a liquid scintillation counter.

20 PAGE analysis was done as above except that 90 pmol of [<sup>14</sup>C]acetyl CoA (55 mCi/mmol, Amersham Life Science Inc.) and 9 pmol of core histones or mononucleosomes were used. Core histones and mononucleosomes were prepared as described (35). For trypsin digestion, reaction mixtures were further incubated with various amounts of trypsin on ice for 30 minutes. The samples were analyzed on one

25 dimensional SDS-PAGE gels or two dimensional gels, where the first dimension was an acid-urea-PAGE gel (44) and the second dimension was an SDS-PAGE gel.

### **Protein expression**

For baculovirus expression, cDNAs corresponding to p300 portions of aa 1-670, aa 671-1194 and aa 1135-2414 were amplified by PCR (EXPAND High Fidelity PCR System; Boehringer Mannheim) as KpnI-NotI fragments. The resulting fragments were

30

subcloned into a baculovirus transfer vector having the FLAG-tag sequence (43). The recombinant viruses were isolated using the BACULOGOLD system (Pharmingen), according to the manufacturer's protocol and were infected into Sf9 cells (ATCC accession number CRL 1711) to express FLAG-p300. Recombinant proteins were  
5 affinity purified with M2 anti-FLAG antibody-immobilized agarose (Kodak-IBI) according to the manufacturer's protocol.

For bacterial expression, cDNAs encoding the p300 portions and the CBP portion (aa 1174-1850) were first subcloned into the baculovirus transfer vector having  
10 the FLAG-tag as described above. Thereafter, the XhoI and NotI fragments encoding FLAG-p300 or FLAG-CBP fusions were resubcloned into the *E. coli* expression vector pET-28c (Novagene) digested with SalI and NotI. Recombinant proteins were expressed in *E. coli* BL21(DE3) and affinity purified with M2-antibody agarose.

#### 15 **Histone acetyltransferases that associate with E1A**

Although the adenovirus E1A 12S protein (E1A) inhibits transcription in a variety of genes via direct binding to p300/CBP (45), E1A also stimulates transcription in some contexts (46). Thus, p300/CBP-bound E1A was tested to determine whether it might recruit histone acetyltransferases or deacetylases to regulate transcription. In  
20 addition, experiments were conducted as described below to determine if p300/CBP per se is a histone acetyltransferase.

Initially, recombinant FLAG-epitope tagged E1A was immobilized on anti-FLAG antibody beads. Immobilized E1A was incubated with a HeLa nuclear  
25 extract for affinity purification of E1A-associated polypeptides. FLAG-E1A was then eluted from the beads, along with E1A-associated polypeptides, by incubating with FLAG-peptide. Although E1A per se has no histone acetyltransferase activity, E1A recruited significant amounts of histone acetyltransferase activity from the nuclear extract. It is very unlikely that this activity is derived from P/CAF given that  
30 E1A and P/CAF cannot bind to p300/CBP simultaneously (43). Consistent with this, no P/CAF was detected in these fractions by immunoblotting.

The E1A N-terminus, a region that is not highly conserved among the various adenovirus serotypes, is involved in p300/CBP binding *in vivo*. Mutations in the N-terminal region lead to loss of the ability for p300/CBP binding without affecting RB binding (1,47). Thus, the requirement of the E1A N-terminal region for the recruitment of histone acetyltransferase activity was tested. In contrast to the wild type, the N-terminal deleted form of E1A ( $\Delta$ N-E1A) recruited only a background level of acetyltransferase activity. In agreement with previous reports (47), the  $\Delta$ N-E1A showed no ability to interact with p300/CBP, although it still retained the ability to interact with a variety of other polypeptides, including RB.

To define the relationship between p300/CBP and histone acetyltransferase activity, affinity purified E1A-binding polypeptides were separated by Mono Q ion-exchange column. Both p300/CBP and the acetyltransferase activity were coeluted at 140 mM KCl, while most of polypeptides were eluted at 260 mM KCl. The active fraction of Mono Q column (~140 mM KCl) was further separated by Superdex-200 gel filtration column. Both p300/CBP and the acetyltransferase activity coeluted after the void volume, indicating that p300/CBP is involved in the histone acetyltransferase activity.

20

#### **p300 is a histone acetyltransferase**

The data provided herein indicate that p300 per se, or a polypeptide(s) associated with p300, possesses histone acetyltransferase activity. To test the former possibility, the acetyltransferase activity of recombinant p300 was measured. p300 was divided into three fragments, each of which was expressed in and purified from Sf9 cells via a baculovirus expression vector. Histone acetyltransferase activity was readily detected in the C-terminal fragment containing amino acids 1135-2414, whereas no activity was found in the other fragments, demonstrating conclusively that p300 per se is a histone acetyltransferase.

30

**p300/CBP-histone acetyltransferase domain**

To map the histone acetyltransferase domain of p300, a series of deletions was prepared. Given the poor conservation of the glutamine-rich region (aa 1815-2414) in the *C. elegans* p300/CBP homolog (6), the p300 fragment encoding aa 1135-1810 was expressed in and purified from *E. coli*. Importantly, this candidate region of p300 (aa 1135-1810) showed significant histone acetyltransferase activity. For further mapping within this region, a series of N-terminal deletions was constructed. Deletion of 60 residues, resulting in a fragment containing aa 1195-1810, had no effect on the acetyltransferase activity, whereas the deletion of 185 residues, yielding a fragment comprising aa residues 1320-1810, completely eliminated the acetyltransferase activity.

Next, a series of C-terminal deletions was analyzed to determine the requirement of the P/CAF (or E1A) -binding domain. The p300 fragments lacking the E1A binding domain (aa 1195-1760, 1195-1706 and 1195-1673) still retained the acetyltransferase activity, whereas the further truncated mutant (aa 1195-1652) completely lost the acetyltransferase activity. Consistent with these results, the internal deletion of residues 1418-1720 showed no acetyltransferase activity. These data demonstrate that the histone acetyltransferase domain is located between the bromodomain and the E1A-binding domain. Given that the histone acetyltransferase domain is highly conserved between p300 and CBP (91% similarity), the corresponding region of CBP, aa residues 1174-1850, was expressed to confirm the acetyltransferase activity. As expected, comparable activity was detected, indicating that both p300 and CBP are histone acetyltransferases.

Among various acetyltransferases including histone acetyltransferases GCN5 and P/CAF, putative acetyl-CoA binding sites are conserved (48). However, multiple alignment analysis (49) showed that the p300/CBP histone acetyltransferase domain does not belong to this group. Moreover, comparison of the p300/CBP histone acetyltransferase domain with peptide sequence databases (23) showed no sequence similarity to any other proteins. Accordingly, this invention shows that p300/CBP represents a novel class of acetyltransferases in that it does not have the conserved motif found among previously described acetyltransferases (48).

**p300 acetylates all core histones in mononucleosomes**

Substrate specificity for acetylation by p300 was also examined. As substrates, histone octamers and mononucleosomes (146 base pairs of DNA wrapped around the octamer of core histones) were used. Given that the histone octamer dissociates into dimers or tetramers under physiological conditions, the histone octamer is referred to here as core histones. When core histones were used, p300 acetylated all four proteins, but preferentially H3 and H4. More importantly, in a nucleosomal context, p300 acetylated all four core histones nearly stoichiometrically. In contrast, p300 acetylated neither BSA nor lysozyme.

Hyperacetylated histones are believed to be linked with transcriptionally active chromatin (26,27,50,51). Hyperacetylated forms are found in histones H4, H3 and H2B, which have multiple acetylation sites *in vivo*. Thus, the level of acetylation by p300 was also tested.

Mononucleosomes treated with p300 were analyzed by two-dimensional gel electrophoresis. A Coomassie blue-stained gel and the corresponding autoradiogram showed that a significant amount of histones, especially H4, were hyperacetylated. Importantly, acetylation levels by p300 were very close to those of hyperacetylated histones prepared from HeLa nuclei treated with sodium butyrate, a histone deacetylase inhibitor. In contrast, no acetylated forms were detected in the reaction without p300. These results indicate that p300 acetylates histones in mononucleosomes to the hyperacetylated state by targeting multiple lysine residues.

**p300 acetylates the four lysines in the histone H4 N-terminal tail *in vitro* which are acetylated *in vivo***

Lysines at positions 5, 8, 12 and 16 of histone H4 are acetylated *in vivo* (51). Recent studies with yeast histone acetyltransferases demonstrate the position-specific acetylation by distinct acetyltransferases, i.e., while cytoplasmic acetyltransferases for histone deposition and chromatin assembly modify

positions 5 and 12, GCN5 modifies positions 8 and 16 (52). Accordingly, the positions of acetylation by p300 were also determined. A series of synthetic peptides containing acetylated lysines at various positions was used to determine the acetylation site-specificity of p300. Consistent with the two-dimensional gel electrophoresis analysis, the experiments with peptide substrates showed that p300 acetylates all four lysines in the histone H4 that are acetylated *in vivo*. These results are consistent with the view that deposition-related diacetylated histones are deacetylated during maturation of chromatin (53).

#### 10 **p300 preferentially acetylates the N-terminal histone tail**

Histone acetyltransferases modify specific lysine residues in the N-terminal tail of core histones but not the C-terminal globular domain *in vivo* (26,27,50,51). Structural models of nucleosomes (54,55,56) suggest that most of the lysine residues in the C-terminal globular domain are buried. Therefore, experiments were conducted to examine whether restricted acetylation of the N-terminal tail resulted from the substrate specificity of the enzyme or inaccessibility of the enzyme to the core domain in nucleosomes. The globular domains of all core histones contain a long helix flanked on either side by a loop segment and short helix, termed the "histone fold" (54,55,56). The histone fold is involved in formation of the stable H2A-H2B and H3-H4 hetero-dimers, consisting of extensive hydrophobic contacts between the paired molecules. Therefore, it is likely that a histone monomer cannot fold properly, thereby increasing access of the histone acetyltransferase to the core domain. Based on these considerations, experiments were conducted to determine whether p300 acetylates free histone H4 in a N-terminal-specific manner.

25

Histone H4 was acetylated with p300 and subsequently the histone tail was removed by partial digestion with trypsin. The distributions of radioactivity between intact and core histones were compared. While the globular core histone domain was predominant at the higher trypsin concentrations, radioactivity was detected mostly in the intact histone. These data demonstrate that p300 preferentially acetylates the N-terminal tail of histone H4.

30

### 5 III. P/CAF interaction with MyoD

#### Tissue culture and transfection experiments

C<sub>2</sub>C<sub>12</sub> mouse cells (ATCC accession number CRL 1772) were grown in Dulbecco's modified Eagle medium (DMEM) supplemented with 20% fetal bovine  
10 serum (FBS) until they reached confluence. Differentiation was induced by switching medium to differentiation medium (DM), consisting of DMEM containing 2% horse serum. C<sub>3</sub>H/10T1/2 fibroblasts (ATCC accession number CCL 226) were grown in DMEM supplemented with 10% FBS. Cells were transfected by the calcium phosphate precipitation method. Total amounts of transfected DNA were equalized by empty  
15 vector DNA. After 12 h incubation in medium containing the precipitated DNA, the cells were washed and incubated in fresh DMEM containing 10% FBS for an additional 24 h. Afterwards, differentiation was induced by incubating in DM for 36 to 72 h. Chloramphenicol acetyltransferase (CAT) assays were performed as previously described (64,69). The quantities of cell extracts used for CAT assays were normalized  
20 to  $\beta$ -galactosidase activity by cotransfection of 1 mg of the  $\beta$ -galactosidase expression vector, pON260.

Expression vectors used for transfection experiments are as follows:

pCX-P/CAF for P/CAF (43); pCMV-bp300 for p300 (65), pCMV-p300 (1869-2414)  
25 (64) and pCMV-p300 (1514-1922) (60) for p300 wild type and mutants; pE1A12S, pE1A12S R2G, pE1A12S D2-36 and pE1A12S D121-130 for E1A wild type and mutants (66,67,68); and pEMSV-MyoD for MyoD (64).

The antisense P/CAF RNA expression vector, pcDNA3 P/CAF-AS, was created  
30 as follows. The 2.5 Kb EcoRI-KpnI fragment containing the entire P/CAF open reading frame was isolated from pCX-P/CAF (43). This fragment was subcloned into the

EcoRI-KpnI sites of plasmid pcDNA3 (Invitrogen) so that the antisense P/CAF RNA is driven under the CMV promoter. Reporter genes employed were 4RE-CAT and MCK-CAT (69). 4RE-CAT is driven by a synthetic promoter containing 4 copies of the E-box, whereas MCK-CAT is driven by the native MCK promoter (nucleotides -1256 to +7).

### Microinjection and immunofluorescence

Cells were grown on small glass slides, subdivided into numbered squares of 2 mm x 2 mm and microinjected with purified and concentrated antibodies, as previously described (70). For immunofluorescence, cells were fixed in either 2% paraformaldehyde or 1:2 methanol/acetone solution, preincubated with 5% BSA/PBS and incubated with the primary antibodies for 30 min at 37° C. Subsequently, antibody was visualized by incubating with either rhodamine- or fluorescein-conjugated secondary antibody for 30 min at 37° C. Injected antibodies were stained with a rhodamine-conjugated secondary antibody and nuclei were counter-stained by DAPI as previously described (69).

Antibodies employed are as follows; rabbit polyclonal affinity purified anti-P/CAF antibody (43), rabbit polyclonal anti-p300/CBP antiserum (71), mouse monoclonal anti-MyoD antibody (clone 5.8A, kindly provided by P. Houghton), goat polyclonal anti-c-Jun affinity purified antibody (Santa Cruz) and rabbit pre-immune serum.

25

### Immunoprecipitation and DNA affinity purification

Cells were resuspended in lysis buffer (20 mM NaPO<sub>4</sub>, 150 mM NaCl, 5mM MgCl<sub>2</sub>, 0.1% NP40, 1 mM DTT, 10 mM sodium fluoride, 0.1 mM sodium vanadate, 1 mM phenylmethylsulfonyl-fluoride and 10 mg/ml each of leupeptin, aprotinin and pepstatin). After 30 min incubation on ice, samples were centrifuged at 12,000 x g for 30 min and supernatants were used as cell extracts. Extracts were pre-cleared by

incubating with rabbit pre-immune serum and protein A/G Plus-Agarose (Santa Cruz) for 2 h at 4 C. For immunoprecipitation, the supernatants were incubated with the respective antibodies for 3 h at 4 C. Protein A/G Plus-Agarose was added, and incubation continued for 3 h. The matrix was washed with lysis buffer, then boiled in 2 X SDS sample buffer. Immunoblotting was performed by using the ECL chemiluminescent detection kit (Amersham) according to the manufacturer's protocol.

Affinity purification of E-box-bound complexes was done as previously described (69). Briefly, 100 ng of the biotinylated double stranded DNA containing the E-box were immobilized on streptavidin-conjugated magnetic beads and incubated with 500 mg of cell extracts in the presence of poly dI-dC. After extensive washing, bound proteins were eluted with SDS sample buffer and analyzed by immunoblotting.

#### **In vitro protein-protein interaction assays**

The CBP-B fragment and its deletion derivatives were expressed as GST-fusions described previously (43). MyoD and E1A (43) were expressed as FLAG-fusion proteins in Sf9 cells via a baculovirus expression system and affinity-purified on M2 anti-FLAG antibody-agarose (Kodak-IBI). Crude E. coli extracts containing GST-fusions were incubated with various amounts of MyoD and/or E1A in 50 ml of buffer B (20 mM Tris-HCl, pH 8.0, 0.1 M KCl, 5 mM MgCl<sub>2</sub>, 10% glycerol, and 0.1% Nonidet P-40) on ice for 10 min. GST-precipitation was performed as described (43). MyoD and E1A were detected by immunoblotting with anti-FLAG M2 antibody. For the interaction between P/CAF and MyoD, 1.5 pmol of FLAG-P/CAF and 15 pmol of FLAG-MyoD were incubated in 50 ml of buffer B on ice for 10 min. The mixture was further incubated with 2 mg of anti-P/CAF (43) or anti-hADA2 antibody for 60 min. The immunocomplexes were precipitated by incubation with 10 ml of protein A-Trisacryl (Pierce) and rotated for 1-4 hr at 4°C. The matrix was washed 4 times with 200 ml of buffer B and boiled in 10 ml of 2 X SDS sample buffer. The proteins were resolved on a 4%-20% gradient SDS-PAGE and subjected to immunoblotting with the anti-FLAG M2 antibody. The blot was developed with the SUPERSIGNAL chemiluminescent substrates (Pierce).

**P/CAF coactivates muscle-specific transcription**

P/CAF and MyoD were co-transfected into mouse C3H10T1/2 fibroblasts, and MyoD-mediated transcription was determined from reporter activity driven by the artificial (4RE) and the naturally-occurring muscle creatine kinase (MCK) promoters. Overexpression of P/CAF stimulated MyoD-dependent transcription several folds in both promoters. Similar results were obtained for the myoD activated myogenin promoter. Transcriptional activation was further stimulated by co-transfecting with MyoD, P/CAF and p300 expression vectors, suggesting that P/CAF may function by forming a complex with p300/CBP. Consistent with the lack of DNA binding capacity in P/CAF, overexpression of P/CAF alone did not increase the basal transcriptional activity of either enhancer. To test whether P/CAF and p300/CBP function in the same pathway, two dominant negative forms of p300 were employed which specifically inhibit p300/CBP-mediated transcription (60,64). The p300 segment spanning residues 1514-1922 inhibits the MyoD-dependent activation via direct interaction with MyoD (60), whereas the p300 segment spanning residues 1869-2414 inhibit it without direct interaction (64). Both dominant negative mutants inhibited MyoD-coactivation by P/CAF), suggesting that P/CAF and p300/CBP function in the same pathway.

For further elucidation of the activation mechanism by P/CAF, the effect of E1A, which inhibits MyoD-dependent transcription and differentiation (66,72,73) via direct interaction with p300/CBP (65,78), was tested. Expression of E1A in C3H10T1/2 fibroblasts inhibited stimulation of MyoD-directed transcription by P/CAF overexpression. E1A mutants lacking p300/CBP-binding activity, E1A D2-36 and E1A R2G (67,79), had almost no effect. On the other hand, an E1A mutant retaining p300/CBP-binding activity, E1A D121-130, behaved like the wild type. Since E1A associates with p300/CBP, but not with P/CAF, these results suggest that P/CAF functions in MyoD-directed transcription via interaction with p300/CBP.

To address the role of P/CAF as a myogenic coactivator in a more relevant environment, P/CAF was overexpressed in proliferating C2C12 myoblasts which express

endogenous myogenic bHLH factors. As observed in fibroblasts, overexpression of P/CAF stimulated muscle specific transcription. Concomitant expression of exogenous p300 increased P/CAF-mediated coactivation. The repression exerted by wild type E1A, but not mutant E1A D2-36, on P/CAF coactivation of MyoD was also observed in muscle cells.

Similar experiments were performed with myogenic cell lines that were stably transformed with wild type or mutant E1A-expressing vectors (66). Coactivation by P/CAF was inhibited by wild type E1A or the E1A mutant that retains p300/CBP-binding activity (E1A $\Delta$ 121-130). In contrast, E1A mutants that lack p300/CBP-binding (E1A  $\Delta$ 2-36 and E1A R2G) allowed transcriptional coactivation by P/CAF. Taken together, these experiments show that P/CAF coactivates MyoD-directed transcription via interaction with p300/CBP.

#### **P/CAF stimulates myogenic differentiation**

Given that P/CAF potentiates MyoD-directed transcription, the ability of P/CAF to assist MyoD in promoting myogenic differentiation was investigated. To this aim, C3H10T1/2 fibroblasts were transiently transfected with P/CAF and MyoD expression vectors. An expression vector for the green fluorescent protein (GFP) was co-transfected to identify transfected cells. After incubation in differentiation medium, the myogenic conversion of transfected cells was determined by simultaneous expression of the GFP and the differentiation-specific marker myosin heavy chain (MHC). Forced expression of MyoD in fibroblasts caused muscle differentiation in 12% of the transfected fibroblasts. This myogenic conversion was 20% by co-expressing MyoD and P/CAF. As observed in transcription experiments, stimulation of differentiation by P/CAF was counteracted by co-transfection with the p300 dominant negative mutant, p300 (1869-2414). Consistent with a general role for coactivators, overexpression of P/CAF alone was unable to differentiate fibroblasts.

Similar experiments were done using proliferating C2C12 myoblasts in which the differentiation program is already committed. Most of the myoblasts differentiated into

myotubes by overexpressing P/CAF, whereas only a modest effect was observed by overexpressing p300. In contrast, differentiation was inhibited slightly by overexpressing c-Jun. This inhibitory effect presumably was caused by titration of p300/CBP, which associates directly with c-Jun (74). A similar inhibition was observed in the p300 dominant negative mutant. Consistent with the transcriptional effect, E1A almost completely inhibited differentiation. The E1A mutant RG2, lacking p300/CBP-binding capability but retaining the retinoblastoma protein (Rb)-binding capability, only partially inhibited differentiation, although this same mutant inhibited transcription as severely as the wild type. Taken together, these data show that P/CAF stimulates muscle differentiation by coactivating MyoD function via p300/CBP association.

#### **P/CAF is essential for myogenic transcription and differentiation**

To test the necessity of P/CAF for myogenic transcription, experiments were conducted whereby P/CAF synthesis was inhibited by expressing antisense P/CAF RNA. A vector from which the P/CAF mRNA is transcribed in the antisense orientation (P/CAF-AS) was transfected with P/CAF and MyoD expression vectors into fibroblasts and MyoD-dependent transcription was examined. Cotransfection of the antisense expression vector strongly inhibited MyoD-dependent transcription below the level of induction elucidated by MyoD alone, demonstrating that expression of P/CAF antisense RNA inhibits not only the coactivation exerted by exogenous P/CAF but also that of endogenous P/CAF. These results indicate that P/CAF is essential for MyoD-dependent transcription.

Studies were also carried out to determine whether expression of P/CAF antisense RNA inhibits myogenic differentiation. C3H10T1/2 fibroblasts were transiently transfected with various expression vectors with or without the P/CAF antisense RNA expression vector. Expression of P/CAF antisense RNA reduced MyoD-mediated myogenic conversion of fibroblasts. Expression of P/CAF antisense RNA also counteracted the stimulatory effect of both P/CAF and p300 on myogenic differentiation. These data support the view that P/CAF and p300/CBP coactivate

MyoD-dependent transcription in the same pathway. More drastic inhibition was observed in C2C12 myoblasts in similar experiments. Therefore, it can be concluded that P/CAF is essential for transcription of muscle specific genes and hence differentiation into myotubes.

5

To further confirm the essential role of P/CAF for myogenic differentiation, we blockage experiments by antibody microinjection were performed. Antibodies were injected into the cytoplasm of proliferating C2C12 myoblasts to prevent the nuclear transport of newly synthesized target proteins. After incubating in the differentiation medium, the degree of differentiation was determined. Microinjection of an anti-P/CAF antibody almost completely inhibited differentiation. Similar results were obtained by microinjecting anti-p300/CBP antibodies. Although microinjection of either anti-p300/CBP or P/CAF antibody was sufficient to inhibit differentiation, an even greater inhibition was observed by coinjecting both of them. Microinjection of anti-P/CAF or anti-p300/CBP antibody did not interfere with induction of p53 by DNA damaging agents, showing specificity of the inhibition by the antibodies. In contrast to anti-P/CAF or anti-p300/CBP antibodies, the injection of anti-MyoD antibody only partially inhibited differentiation, supporting the view of functional redundancy between MyoD and Myf-5 (75,76). Injection of anti-c-Jun antibody or control antibody did not interfere with muscle differentiation.

15  
20

Similar experiments were performed with C3H10T1/2 fibroblasts stably expressing MyoD. In these cells, either anti-p300/CBP or anti-P/CAF antibody completely inhibited muscle differentiation. In contrast to myoblasts, anti-MyoD antibody completely blocked differentiation in the fibroblasts expressing MyoD. Anti-c-Jun and control antibodies did not interfere with differentiation. Taken together, these results demonstrate that P/CAF and p300/CBP are indispensable for activation of the myogenic program.

25

30

**p300/CBP, P/CAF and MyoD form a multimeric complex *in vivo***

The data described above indicate that P/CAF stimulates MyoD-directed transcription via association with p300/CBP. Thus, experiments were conducted to investigate whether P/CAF, p300/CBP and MyoD could associate in a complex.

- 5 First, cellular extracts derived from C2C12 myotubes were subjected to immunoprecipitation. Both anti-MyoD and anti-p300/CBP antibodies co-precipitated P/CAF. In a complementary experiment, both anti-p300/CBP and anti-P/CAF antibodies also co-precipitated MyoD, suggesting that these factors form a multimeric protein complex in myotubes.

10.

- Next, attempts were made to detect this complex on the E-box, the DNA binding site for MyoD. Immobilized DNA containing an E-box sequence was incubated with myotube extracts. After extensive washing, P/CAF, p300/CBP and MyoD were analyzed by immunoblotting. P/CAF, p300/CBP and MyoD were all affinity purified on  
15 the immobilized DNA, whereas they were not purified on the control DNA lacking the E-box. Given that P/CAF and p300/CBP per se cannot bind to DNA, these observations indicate that P/CAF and p300/CBP are recruited through MyoD at the E-box sites to form a multi-protein complex.

**20 Complex formation is inhibited by viral transforming factors**

- Since the oncoviral proteins E1A and large T antigen inhibit myogenic transcription and differentiation, the effect of these factors on the formation of complexes on the E-box was tested. Importantly, very small amounts of P/CAF and p300/CBP were co-purified on the E-box from myocyte extracts which stably express  
25 E1A or large T antigen, although MyoD was detected under these conditions. The lower recovery of MyoD from E1A-expressing muscle cells could reflect the low level of MyoD in the extracts (66). These results indicate that E1A and large T antigen dissociate P/CAF and p300/CBP from MyoD without altering MyoD binding to DNA.

- 30 Consistent with the previous observations that transiently expressed E1A prevents interaction between P/CAF and p300/CBP *in vivo* (43), the association

between p300/CBP and P/CAF was abolished in myoblasts stably transformed by wild type E1A but not in those clones transformed with the E1A mutant R2G unable to bind p300/CBP. Similarly, the interaction between p300/CBP and P/CAF was abolished by large T antigen but not by the mutant protein that localizes into the cytoplasm (77).

5

#### **Interaction between MyoD, P/CAF and CBP *in vitro***

Previous interaction experiments *in vitro* indicate that the CBP region spanning residues 1801 to 1850 is crucial for interaction with both P/CAF and E1A (43). While most sequence-specific factors bind to CBP sites distinct from the P/CAF/E1A binding sites, MyoD interacts with an overlapping CBP fragment called the CH3 region (60,64,65). To understand how P/CAF, p300/CBP and MyoD associate, the CBP sites important for MyoD binding were mapped more precisely. Consistent with previous reports (60,64,65), the CBP fragment spanning residues 1801-2000 (fragment B) bound MyoD. Moreover, deletion of residues 1801 to 1850 within fragment B completely abolished interaction with MyoD, which is similar to the results obtained with P/CAF and E1A. Importantly, an internal deletion of residues 1850-1878 abolished the MyoD interaction with CBP, while it did not affect binding of E1A or P/CAF (43). These results suggest that MyoD and P/CAF bind to distinct sites of p300/CBP, albeit the binding sites may overlap. Moreover, a direct interaction was observed between MyoD and P/CAF, which may contribute to stabilization of the multimeric complex.

These data show that E1A prevents not only p300/CBP-interaction with P/CAF but also that with MyoD *in vivo*. To obtain evidence that this inhibition is due to the direct action by E1A, competition experiments were performed *in vitro*. Importantly, the interaction between CBP and MyoD was strongly inhibited by addition of E1A, implicating that E1A inhibits myogenic transcription by disrupting multiple interactions.

Although the present process has been described with reference to specific details of certain embodiments thereof, it is not intended that such details should be

30

regarded as limitations upon the scope of the invention except as and to the extent that they are included in the accompanying claims.

Throughout this application various publications are referenced by numbers within parentheses. Full citations for these publications are as follows. The disclosures of these publications in their entireties are hereby incorporated by reference into this application in order to more fully describe the state of the art to which this invention pertains.

### References

1. Moran, E. *Curr. Opin. Genet. Dev.* **3**, 63-70 (1993).
2. Dyson, N. & Harlow, E. *Cancer Surv.* **12**, 161-195 (1992).
3. Chrivia, J. C., *et al.* *Nature* **365**, 855-859 (1993).
4. Kwok, R. P., *et al.* *Nature* **370**, 223-226 (1994).
5. Eckner, R., *et al.* *Genes Dev.* **8**, 869-884 (1994).
6. Arany, Z., Sellers, W. R., Livingston, D. M. & Eckner, R. *Cell* **77**, 799-800 (1994).
7. Arany, Z., Newsome, D., Oldread, E., Livingston, D. M. & Eckner, R. *Nature* **374**, 81-84 (1995).
8. Lundblad, J. R., Kwok, R. P., Lurance, M. E., Harter, M. L. & Goodman, R. H. *Nature* **374**, 85-88 (1995).
9. Arias, J., *et al.* *Nature* **370**, 226-229 (1994).
10. Lee, J. S., *et al.* *Genes Dev.* **9**, 1188-1198 (1995).
11. Banister, A. J. & T., K. *EMBO J.* **14**, 4758-4762 (1995).
12. Dai, P. *et al.* *Genes Dev.* **10**, 528-540 (1996).
13. Kamei, Y. *et al.* *Cell* **85**, 403-414 (1996).
14. Rochette-Egly, C., Fromental, C., and Chambon, P. *Genes Dev.* **5**, 1200-1211 (1990).
15. Struhl, K. *Cell* **50**, 841-846 (1987).

16. Engelberg, D., Klein, C., Martinetto, H., Struhl, K. & Karin, M. *Cell* **77**, 381-390 (1994).
17. Georgakopoulos, T. & Thireos, G. *EMBO J.* **11**, 4145-4152 (1992).
18. Berger, S. L., *et al.* *Cell* **70**, 251-265 (1992).
19. Horiuchi, J., Silverman, N., Marcus, G. A. & Guarente, L. *Mol. Cell. Biol.* **15**, 1203-1209 (1995).
20. Berger, S. L., Cress, W. D., Cress, A., Triezenberg, S. J. & Guarente, L. *Cell* **61**, 1199-1208 (1990).
21. Marcus, G. A., Silverman, N., Berger, S. L., Horiuchi, J. & Guarente, L. *EMBO J.* **13**, 4807-4815 (1994).
22. Georgakopoulos, T., Gounalaki, N. & Thireos, G. *Mol. Gen. Genet.* **246**, 723-728 (1995).
23. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. *J. Mol. Biol.* **215**, 403-410 (1990).
24. Adams, M. D., Soares, M. B., Kerlavage, A. R., Fields, C. & Venter, J. C. *Nature Genet.* **4**, 373-380 (1993).
25. Candau, R., *et al.* *Mol. Cell. Biol.* **16**, 593-602 (1996).
26. Turner, B. M. *Cell* **75**, 5-8 (1993).
27. Wolffe, A. P. *Trends Biochem. Sci.* **19**, 240-244 (1994).
28. Brownell, J. E., *et al.* *Cell* **84**, 843-851 (1996).
29. Haynes, S. R., *et al.* *Nucl. Acids Res.* **20**, 2063 (1992).
30. Hopp, T. P., *et al.* *Biotechnology* **6**, 1205-1210 (1988).
31. Giordano, T., Howard, T. H., Coleman, J., Sakamoto, K. & Howard, B. H. *Exp. Cell Res.* **192**, 193-197 (1991).
32. Ogryzko, V. V., Hirai, T. H., Shih, C. E. & Howard, B. H. *J. Virol.* **68**, 3724-3732 (1994).
33. Harlow, E. D. & Lane, D. *Antibodies: A laboratory manual* (Cold Spring Harbor Laboratory Press, New York, 1988).
34. Minucci, S., *et al.* *Mol. Cell. Biol.* **14**, 360-372 (1994).
35. Hayes, J. J., Pruss, D. & Wolffe, A. P. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 7817-7821 (1994).

36. Brownell, J. E. & Allis, C. D. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 6364-6368 (1995).
37. Dayhoff et al. *Atlas of Protein Sequence and Structure* 1978, Nat'l Biomed. Res. Found., Washington, D.C.
38. Pei et al., *J. Biol. Chem.* **266**:16363-16369 (1991).
39. Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989.
40. Cunningham, et al. *Science* **243**:1330-1336 (1989).
41. Ferretti, et al. *Proc. Nat. Acad. Sci.* **82**:599-603 (1986).
42. Brake et al. *PNAS* **82**:4642-4646 (1984).
43. Yang, X.J., et al. *Nature* **382**:319-324 (1996).
44. Panyim, S. and Chalkley, R. *Arch. Biochem. Biophys.* **130**:337-346 (1969).
45. Song, C.Z., et al. *J. Biol. Chem.* **264**:24524-24530 (1989).
46. Kitabayashi, I., et al. *EMBO J.* **14**:3496-3509 (1995).
47. Wang, H.G., et al. *J. Virol.* **67**:4804-4813 (1993).
48. Lu, L., et al. RGF GIGS is an amino acid sequence required for acetyl coenzyme A binding and activity of human spermidine/spermine N1 acetyltransferase. **271**:18920-18924 (1996).
49. Lawrence, C.E., et al. *Science* **262**:208-214 (1993).
50. Wolffe, A.P. and Pruss, D. *Cell* **86**:817-819 (1996).
51. Brownell, J.E. and Allis, C.D. *Curr. Opin. Genet. Dev.* **6**:176-184 (1996).
52. Kuo, M.-H., et al. *Nature* **383**:269-272 (1996).
53. Shimamura, A. and Worcel, A. *J. Biol. Chem.* **264**:14524-14530 (1989).
54. Arents, G. and Moudrianakis, E.N. *PNAS* **90**:10489-10493 (1993).
55. Ramakrishnan, V. *Curr. Opin. Struct. Biol.* **4**:44-50 (1994).
56. Pruss, D., et al. *Bioassays* **17**:161-170 (1995).
57. Chakravarti et al. *Nature* **383**:99-103 (1996).
58. Oelgeschlager et al. *EMBO J.* **15**:2771-2780 (1996).
59. Janknecht and Nordheim *Oncogene* **12**:1961-1969 (1996).
60. Yuan, W., Condorelli, G., Caruso, M., Felsani, A., and Giordano, A. Human p300 protein is a coactivator for the transcription factor MyoD. *J.*

- Biol. Chem.* **271**:9009-9013 (1996).
61. Onate et al. *Science* **270**:1354-1357 (1995).
  62. A. Melton, Ed., *Antisense RNA and DNA*, D. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1988).
  63. Puri, P. L., Avantaggiati, M. L., Balsano, C., Sang, N., Graessmann, A., Giordano, A., and Levrero, M. p300 is required for MyoD-dependent cell cycle arrest and muscle-specific gene transcription. *EMBO J.* **16**:369-383 (1997).
  64. Sartorelli, V., Huang, J., Hamamori, Y., and Kedes, L. Molecular mechanisms of myogenic coactivation by p300: direct interaction with the activation domain of MyoD and with the MADS box of MEF2C. *Mol. Cell. Biol.* **17**:1010-1026 (1997).
  65. Eckner, R., Yao, T. P., Oldread, E., and Livingston, D. M. Interaction and functional collaboration of p300/CBP and bHLH factors in muscle and B-cell differentiation. *Genes Dev.* **10**:2478-2490 (1996b).
  66. Caruso, M., Martelli, F., Giordano, A., and Felsani, A. Regulation of MyoD gene transcription and protein function by the transforming domains of the adenovirus E1A oncoprotein. *Oncogene* **8**:267-278 (1993).
  67. Wang, H. G., Rikitake, Y., Carter, M. C., Yaciuk, P., Abraham, S. E., Zerler, B., and Moran, E. Identification of specific adenovirus E1A N-terminal residues critical to the binding of cellular proteins and to the control of cell growth. *J. Virol.* **67**:476-488 (1993a).
  68. Wang, H. G., Yaciuk, P., Ricciardi, R. P., Green, M., Yokoyama, K., and Moran, E. The E1A products of oncogenic adenovirus serotype 12 include amino-terminally modified forms able to bind the retinoblastoma protein but not p300. *J. Virol.* **67**:4804-4013 (1993b).
  69. Puri, P. L., Avantaggiati, M. L., Balsano, C., Sang, N., Graessmann, A., Giordano, A., and Levrero, M. p300 is required for MyoD-dependent cell cycle arrest and muscle-specific gene transcription. *EMBO J.* **16**:369-383 (1997).
  70. Graessmann, M. and Graessmann, A. Microinjection of tissue culture

- cells. *Methods Enzymol.* **101**:482-492 (1983).
71. Avantaggiati, M. L., Carbone, M., Graessmann, A., Nakatani, Y., Howard, B., and Levine, A. S. The SV40 large T antigen and adenovirus E1a oncoproteins interact with distinct isoforms of the transcriptional co-activator, p300. *EMBO J.* **15**:2236-2248 (1996).
72. Webster, K. A., Muscat, G. E., and Kedes, L. Adenovirus E1A products suppress myogenic differentiation and inhibit transcription from muscle-specific promoters. *Nature* **332**:553-557 (1988).
73. Mymryk, J. S., Lee, R. W., and Bayley, S. T. Ability of adenovirus 5 E1A proteins to suppress differentiation of BC3H1 myoblasts correlates with their binding to a 300 kDa cellular protein. *Mol. Biol. Cell.* **3**:1107-1115 (1992).
74. Arias, J., Alberts, A. S., Brindle, P., Claret, F. X., Smeal, T., Karin, M., Feramisco, J., and Montminy, M. Activation of cAMP and mitogen responsive genes relies on a common nuclear factor. *Nature* **370**:226-229 (1994).
75. Rudnicki, M. A., Braun, T., Hinuma, S., and Jaenisch, R. Inactivation of MyoD in mice leads to up-regulation of the myogenic HLH gene Myf-5 and results in apparently normal muscle development. *Cell* **71**, 383-390 (1992).
76. Rudnicki, M. A., Schnegelsberg, P. N., Stead, R. H., Braun, T., Arnold, H. H., and Jaenisch, R. MyoD or Myf-5 is required for the formation of skeletal muscle. *Cell* **75**:1351-1359 (1993).
77. Tedesco, D., Caruso, M., Fischer, F. L., and Vesco, C. The inhibition of cultured myoblast differentiation by the simian virus 40 large T antigen occurs after myogenin expression and Rb up-regulation and is not exerted by transformation-competent cytoplasmic mutants. *J. Virol.* **69**:6947-6957 (1995).
78. Moran, E. DNA tumor virus transforming proteins and the cell cycle. *Curr. Opin. Genet. Dev.* **3**:63-70 (1993).
79. Giordano et al. *Exp. Cell. Res.* **192**:193-197 (1991).
80. Brownell and Allis *PNAS* **92**:6364-6368 (1995).

## SEQUENCE LISTING

## (1) GENERAL INFORMATION

- (i) APPLICANT: The United States of America, as represented by the Secretary, Department of Health and Human Services, c/o National Institutes of Health, Office of Technology Transfer, 6011 Executive Boulevard, Suite 325, Rockville, Maryland 20842
- (ii) TITLE OF THE INVENTION: METHODS AND COMPOSITIONS FOR p300/CBP-ASSOCIATED TRANSCRIPTIONAL CO-FACTOR P/CAF
- (iii) NUMBER OF SEQUENCES: 18
- (iv) CORRESPONDENCE ADDRESS:
- (A) ADDRESSEE: NEEDLE & ROSENBERG, P.C.
  - (B) STREET: Suite 1200, 127 Peachtree Street, NE
  - (C) CITY: Atlanta
  - (D) STATE: GA
  - (E) COUNTRY: USA
  - (F) ZIP: 30303
- (v) COMPUTER READABLE FORM:
- (A) MEDIUM TYPE: Diskette
  - (B) COMPUTER: IBM Compatible
  - (C) OPERATING SYSTEM: DOS
  - (D) SOFTWARE: FastSEQ for Windows Version 2.0
- (vi) CURRENT APPLICATION DATA:
- (A) APPLICATION NUMBER:
  - (B) FILING DATE: 23-JUL-1997
  - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
- (A) APPLICATION NUMBER: Corresponding U.S. Serial No. 60/022,273
  - (B) FILING DATE: 23-July-1996
- (viii) ATTORNEY/AGENT INFORMATION:
- (A) NAME: Miller, Mary L
  - (B) REGISTRATION NUMBER: 39,303
  - (C) REFERENCE/DOCKET NUMBER: 14014.0238/P
- (ix) TELECOMMUNICATION INFORMATION:
- (A) TELEPHONE: 404/688-0770
  - (B) TELEFAX: 404/688-9880
  - (C) TELEX:

## (2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 832 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: None

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Met	Ser	Glu	Ala	Gly	Gly	Ala	Gly	Pro	Gly	Gly	Cys	Gly	Ala	Gly	Ala	1	5	10	15
Gly	Ala	Gly	Ala	Gly	Pro	Gly	Ala	Leu	Pro	Pro	Gln	Pro	Ala	Ala	Leu	20	25	30	
Pro	Pro	Ala	Pro	Pro	Gln	Gly	Ser	Pro	Cys	Ala	Ala	Ala	Ala	Gly	Gly	35	40	45	
Ser	Gly	Ala	Cys	Gly	Pro	Ala	Thr	Ala	Val	Ala	Ala	Ala	Gly	Thr	Ala	50	55	60	
Glu	Gly	Pro	Gly	Gly	Gly	Gly	Ser	Ala	Arg	Ile	Ala	Val	Lys	Lys	Ala	65	70	75	80
Gln	Leu	Arg	Ser	Ala	Pro	Arg	Ala	Lys	Lys	Leu	Glu	Lys	Leu	Gly	Val	85	90	95	
Tyr	Ser	Ala	Cys	Lys	Ala	Glu	Glu	Ser	Cys	Lys	Cys	Asn	Gly	Trp	Lys	100	105	110	
Asn	Pro	Asn	Pro	Ser	Pro	Thr	Pro	Pro	Arg	Ala	Asp	Leu	Gln	Gln	Ile	115	120	125	
Ile	Val	Ser	Leu	Thr	Glu	Ser	Cys	Arg	Ser	Cys	Ser	His	Ala	Leu	Ala	130	135	140	
Ala	His	Val	Ser	His	Leu	Glu	Asn	Val	Ser	Glu	Glu	Glu	Met	Asn	Arg	145	150	155	160
Leu	Leu	Gly	Ile	Val	Leu	Asp	Val	Glu	Tyr	Leu	Phe	Thr	Cys	Val	His	165	170	175	
Lys	Glu	Glu	Asp	Ala	Asp	Thr	Lys	Gln	Val	Tyr	Phe	Tyr	Leu	Phe	Lys	180	185	190	
Leu	Leu	Arg	Lys	Ser	Ile	Leu	Gln	Arg	Gly	Lys	Pro	Val	Val	Glu	Gly	195	200	205	
Ser	Leu	Glu	Lys	Lys	Pro	Pro	Phe	Glu	Lys	Pro	Ser	Ile	Glu	Gln	Gly	210	215	220	
Val	Asn	Asn	Phe	Val	Gln	Tyr	Lys	Phe	Ser	His	Leu	Pro	Ala	Lys	Glu	225	230	235	240
Arg	Gln	Thr	Ile	Val	Glu	Leu	Ala	Lys	Met	Phe	Leu	Asn	Arg	Ile	Asn	245	250	255	
Tyr	Trp	His	Leu	Glu	Ala	Pro	Ser	Gln	Arg	Arg	Leu	Arg	Ser	Pro	Asn	260	265	270	
Asp	Asp	Ile	Ser	Gly	Tyr	Lys	Glu	Asn	Tyr	Thr	Arg	Trp	Leu	Cys	Tyr	275	280	285	
Cys	Asn	Val	Pro	Gln	Phe	Cys	Asp	Ser	Leu	Pro	Arg	Tyr	Glu	Thr	Thr	290	295	300	
Gln	Val	Phe	Gly	Arg	Thr	Leu	Leu	Arg	Ser	Val	Phe	Thr	Val	Met	Arg	305	310	315	320
Arg	Gln	Leu	Leu	Glu	Gln	Ala	Arg	Gln	Glu	Lys	Asp	Lys	Leu	Pro	Leu	325	330	335	
Glu	Lys	Arg	Thr	Leu	Ile	Leu	Thr	His	Phe	Pro	Lys	Phe	Leu	Ser	Met	340	345	350	
Leu	Glu	Glu	Glu	Val	Tyr	Ser	Gln	Asn	Ser	Pro	Ile	Trp	Asp	Gln	Asp	355	360	365	
Phe	Leu	Ser	Ala	Ser	Ser	Arg	Thr	Ser	Gln	Leu	Gly	Ile	Gln	Thr	Val	370	375	380	
Ile	Asn	Pro	Pro	Pro	Val	Ala	Gly	Thr	Ile	Ser	Tyr	Asn	Ser	Thr	Ser	385	390	395	400
Ser	Ser	Leu	Glu	Gln	Pro	Asn	Ala	Gly	Ser	Ser	Ser	Pro	Ala	Cys	Lys	405	410	415	
Ala	Ser	Ser	Gly	Leu	Glu	Ala	Asn	Pro	Gly	Glu	Lys	Arg	Lys	Met	Thr	420	425	430	
Asp	Ser	His	Val	Leu	Glu	Glu	Ala	Lys	Lys	Pro	Arg	Val	Met	Gly	Asp	435	440	445	
Ile	Pro	Met	Glu	Leu	Ile	Asn	Glu	Val	Met	Ser	Thr	Ile	Thr	Asp	Pro	450	455	460	
Ala	Ala	Met	Leu	Gly	Pro	Glu	Thr	Asn	Phe	Leu	Ser	Ala	His	Ser	Ala	465	470	475	480

Arg	Asp	Glu	Ala	Ala	Arg	Leu	Glu	Glu	Arg	Arg	Gly	Val	Ile	Glu	Phe
				485					490					495	
His	Val	Val	Gly	Asn	Ser	Leu	Asn	Gln	Lys	Pro	Asn	Lys	Lys	Ile	Leu
			500					505					510		
Met	Trp	Leu	Val	Gly	Leu	Gln	Asn	Val	Phe	Ser	His	Gln	Leu	Pro	Arg
		515					520					525			
Met	Pro	Lys	Glu	Tyr	Ile	Thr	Arg	Leu	Val	Phe	Asp	Pro	Lys	His	Lys
	530					535					540				
Thr	Leu	Ala	Leu	Ile	Lys	Asp	Gly	Arg	Val	Ile	Gly	Gly	Ile	Cys	Phe
545					550					555					560
Arg	Met	Phe	Pro	Ser	Gln	Gly	Phe	Thr	Glu	Ile	Val	Phe	Cys	Ala	Val
				565					570					575	
Thr	Ser	Asn	Glu	Gln	Val	Lys	Gly	Tyr	Gly	Thr	His	Leu	Met	Asn	His
			580					585					590		
Leu	Lys	Glu	Tyr	His	Ile	Lys	His	Asp	Ile	Leu	Asn	Phe	Leu	Thr	Tyr
		595					600					605			
Ala	Asp	Glu	Tyr	Ala	Ile	Gly	Tyr	Phe	Lys	Lys	Gln	Gly	Phe	Ser	Lys
	610					615					620				
Glu	Ile	Lys	Ile	Pro	Lys	Thr	Lys	Tyr	Val	Gly	Tyr	Ile	Lys	Asp	Tyr
625					630					635					640
Glu	Gly	Ala	Thr	Leu	Met	Gly	Cys	Glu	Leu	Asn	Pro	Arg	Ile	Pro	Tyr
				645					650					655	
Thr	Glu	Phe	Ser	Val	Ile	Ile	Lys	Lys	Gln	Lys	Glu	Ile	Ile	Lys	Lys
			660					665					670		
Leu	Ile	Glu	Arg	Lys	Gln	Ala	Gln	Ile	Arg	Lys	Val	Tyr	Pro	Gly	Leu
		675					680					685			
Ser	Cys	Phe	Lys	Asp	Gly	Val	Arg	Gln	Ile	Pro	Ile	Glu	Ser	Ile	Pro
	690					695					700				
Gly	Ile	Arg	Glu	Thr	Gly	Trp	Lys	Pro	Ser	Gly	Lys	Glu	Lys	Ser	Lys
705					710					715					720
Glu	Pro	Arg	Asp	Pro	Asp	Gln	Leu	Tyr	Ser	Thr	Leu	Lys	Ser	Ile	Leu
				725					730					735	
Gln	Gln	Val	Lys	Ser	His	Gln	Ser	Ala	Trp	Pro	Phe	Met	Glu	Pro	Val
			740					745					750		
Lys	Arg	Thr	Glu	Ala	Pro	Gly	Tyr	Tyr	Glu	Val	Ile	Arg	Ser	Pro	Met
		755					760					765			
Asp	Leu	Lys	Thr	Met	Ser	Glu	Arg	Leu	Lys	Asn	Arg	Tyr	Tyr	Val	Ser
	770					775					780				
Lys	Lys	Leu	Phe	Met	Ala	Asp	Leu	Gln	Arg	Val	Phe	Thr	Asn	Cys	Lys
785					790					795					800
Glu	Tyr	Asn	Ala	Pro	Glu	Ser	Glu	Tyr	Tyr	Lys	Cys	Ala	Asn	Ile	Leu
			805						810					815	
Glu	Lys	Phe	Phe	Ser	Lys	Ile	Lys	Glu	Ala	Gly	Leu	Ile	Asp	Lys	
			820				825								

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 481 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met	Leu	Glu	Glu	Glu	Val	Tyr	Ser	Gln	Asn	Ser	Pro	Ile	Trp	Asp	Gln
1				5					10					15	
Asp	Phe	Leu	Ser	Ala	Ser	Ser	Arg	Thr	Ser	Gln	Leu	Gly	Ile	Gln	Thr
			20					25					30		
Val	Ile	Asn	Pro	Pro	Pro	Val	Ala	Gly	Thr	Ile	Ser	Tyr	Asn	Ser	Thr
		35					40					45			

Ser	Ser	Ser	Leu	Glu	Gln	Pro	Asn	Ala	Gly	Ser	Ser	Ser	Pro	Ala	Cys
50						55				60					
Lys	Ala	Ser	Ser	Gly	Leu	Glu	Ala	Asn	Pro	Gly	Glu	Lys	Arg	Lys	Met
65				70					75						80
Thr	Asp	Ser	His	Val	Leu	Glu	Glu	Ala	Lys	Lys	Pro	Arg	Val	Met	Gly
			85						90					95	
Asp	Ile	Pro	Met	Glu	Leu	Ile	Asn	Glu	Val	Met	Ser	Thr	Ile	Thr	Asp
			100					105					110		
Pro	Ala	Ala	Met	Leu	Gly	Pro	Glu	Thr	Asn	Phe	Leu	Ser	Ala	His	Ser
		115					120					125			
Ala	Arg	Asp	Glu	Ala	Ala	Arg	Leu	Glu	Glu	Arg	Arg	Gly	Val	Ile	Glu
	130					135					140				
Phe	His	Val	Val	Gly	Asn	Ser	Leu	Asn	Gln	Lys	Pro	Asn	Lys	Lys	Ile
145				150						155					160
Leu	Met	Trp	Leu	Val	Gly	Leu	Gln	Asn	Val	Phe	Ser	His	Gln	Leu	Pro
			165						170					175	
Arg	Met	Pro	Lys	Glu	Tyr	Ile	Thr	Arg	Leu	Val	Phe	Asp	Pro	Lys	His
			180					185					190		
Lys	Thr	Leu	Ala	Leu	Ile	Lys	Asp	Gly	Arg	Val	Ile	Gly	Gly	Ile	Cys
	195						200					205			
Phe	Arg	Met	Phe	Pro	Ser	Gln	Gly	Phe	Thr	Glu	Ile	Val	Phe	Cys	Ala
	210					215					220				
Val	Thr	Ser	Asn	Glu	Gln	Val	Lys	Gly	Tyr	Gly	Thr	His	Leu	Met	Asn
225				230						235					240
His	Leu	Lys	Glu	Tyr	His	Ile	Lys	His	Asp	Ile	Leu	Asn	Phe	Leu	Thr
			245						250					255	
Tyr	Ala	Asp	Glu	Tyr	Ala	Ile	Gly	Tyr	Phe	Lys	Lys	Gln	Gly	Phe	Ser
			260					265					270		
Lys	Glu	Ile	Lys	Ile	Pro	Lys	Thr	Lys	Tyr	Val	Gly	Tyr	Ile	Lys	Asp
	275						280					285			
Tyr	Glu	Gly	Ala	Thr	Leu	Met	Gly	Cys	Glu	Leu	Asn	Pro	Arg	Ile	Pro
	290					295					300				
Tyr	Thr	Glu	Phe	Ser	Val	Ile	Ile	Lys	Lys	Gln	Lys	Glu	Ile	Ile	Lys
305					310					315					320
Lys	Leu	Ile	Glu	Arg	Lys	Gln	Ala	Gln	Ile	Arg	Lys	Val	Tyr	Pro	Gly
			325						330					335	
Leu	Ser	Cys	Phe	Lys	Asp	Gly	Val	Arg	Gln	Ile	Pro	Ile	Glu	Ser	Ile
			340					345					350		
Pro	Gly	Ile	Arg	Glu	Thr	Gly	Trp	Lys	Pro	Ser	Gly	Lys	Glu	Lys	Ser
		355					360					365			
Lys	Glu	Pro	Arg	Asp	Pro	Asp	Gln	Leu	Tyr	Ser	Thr	Leu	Lys	Ser	Ile
	370					375					380				
Leu	Gln	Gln	Val	Lys	Ser	His	Gln	Ser	Ala	Trp	Pro	Phe	Met	Glu	Pro
385						390				395					400
Val	Lys	Arg	Thr	Glu	Ala	Pro	Gly	Tyr	Tyr	Glu	Val	Ile	Arg	Ser	Pro
			405						410					415	
Met	Asp	Leu	Lys	Thr	Met	Ser	Glu	Arg	Leu	Lys	Asn	Arg	Tyr	Tyr	Val
			420					425					430		
Ser	Lys	Lys	Leu	Phe	Met	Ala	Asp	Leu	Gln	Arg	Val	Phe	Thr	Asn	Cys
		435					440					445			
Lys	Glu	Tyr	Asn	Ala	Pro	Glu	Ser	Glu	Tyr	Tyr	Lys	Cys	Ala	Asn	Ile
	450					455					460				
Leu	Glu	Lys	Phe	Phe	Phe	Ser	Lys	Ile	Lys	Glu	Ala	Gly	Leu	Ile	Asp
465					470					475					480
Lys															

## (2) INFORMATION FOR SEQ ID NO:3:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 203 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```

Arg Val Val Gln His Thr Lys Gly Cys Lys Arg Lys Thr Asn Gly Gly
 1          5          10          15
Cys Pro Ile Cys Lys Gln Leu Ile Ala Leu Cys Cys Tyr His Ala Lys
 20          25          30
His Cys Gln Glu Asn Lys Cys Pro Val Pro Phe Cys Leu Asn Ile Lys
 35          40          45
Gln Lys Leu Arg Gln Gln Gln Leu Gln His Arg Leu Gln Gln Ala Gln
 50          55          60
Met Leu Arg Arg Arg Met Ala Ser Met Arg Thr Gly Val Val Gly Gln
 65          70          75          80
Gln Gln Gly Leu Pro Ser Pro Thr Pro Ala Thr Pro Thr Thr Pro Thr
 85          90          95
Gly Gln Gln Pro Thr Thr Pro Gln Thr Pro Gln Pro Thr Ser Gln Pro
100          105          110
Gln Pro Thr Pro Pro Asn Ser Met Pro Pro Tyr Leu Pro Arg Thr Gln
115          120          125
Ala Ala Gly Pro Val Ser Gln Gly Lys Ala Ala Gly Gln Val Thr Pro
130          135          140
Pro Thr Pro Pro Gln Thr Ala Gln Pro Pro Leu Pro Gly Pro Pro Pro
145          150          155          160
Thr Ala Val Glu Met Ala Met Gln Ile Gln Arg Ala Ala Glu Thr Gln
165          170          175
Arg Gln Met Ala His Val Gln Ile Phe Gln Arg Pro Ile Gln His Gln
180          185          190
Met Pro Pro Met Thr Pro Met Ala Pro Met Gly
195          200

```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 351 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```

Met Ser Glu Ala Gly Gly Ala Gly Pro Gly Gly Cys Gly Ala Gly Ala
 1          5          10          15
Gly Ala Gly Ala Gly Pro Gly Ala Leu Pro Pro Gln Pro Ala Ala Leu
 20          25          30
Pro Pro Ala Pro Pro Gln Gly Ser Pro Cys Ala Ala Ala Gly Gly
 35          40          45
Ser Gly Ala Cys Gly Pro Ala Thr Ala Val Ala Ala Gly Thr Ala
 50          55          60
Glu Gly Pro Gly Gly Gly Gly Ser Ala Arg Ile Ala Val Lys Lys Ala
 65          70          75          80
Gln Leu Arg Ser Ala Pro Arg Ala Lys Lys Leu Glu Lys Leu Gly Val
 85          90          95
Tyr Ser Ala Cys Lys Ala Glu Glu Ser Cys Lys Cys Asn Gly Trp Lys
100          105          110
Asn Pro Asn Pro Ser Pro Thr Pro Pro Arg Ala Asp Leu Gln Gln Ile
115          120          125
Ile Val Ser Leu Thr Glu Ser Cys Arg Ser Cys Ser His Ala Leu Ala
130          135          140

```

Ala	His	Val	Ser	His	Leu	Glu	Asn	Val	Ser	Glu	Glu	Glu	Met	Asn	Arg
145					150					155					160
Leu	Leu	Gly	Ile	Val	Leu	Asp	Val	Glu	Tyr	Leu	Phe	Thr	Cys	Val	His
				165					170					175	
Lys	Glu	Glu	Asp	Ala	Asp	Thr	Lys	Gln	Val	Tyr	Phe	Tyr	Leu	Phe	Lys
			180					185					190		
Leu	Leu	Arg	Lys	Ser	Ile	Leu	Gln	Arg	Gly	Lys	Pro	Val	Val	Glu	Gly
		195					200					205			
Ser	Leu	Glu	Lys	Lys	Pro	Pro	Phe	Glu	Lys	Pro	Ser	Ile	Glu	Gln	Gly
	210					215					220				
Val	Asn	Asn	Phe	Val	Gln	Tyr	Lys	Phe	Ser	His	Leu	Pro	Ala	Lys	Glu
225					230					235					240
Arg	Gln	Thr	Ile	Val	Glu	Leu	Ala	Lys	Met	Phe	Leu	Asn	Arg	Ile	Asn
				245					250					255	
Tyr	Trp	His	Leu	Glu	Ala	Pro	Ser	Gln	Arg	Arg	Leu	Arg	Ser	Pro	Asn
			260					265					270		
Asp	Asp	Ile	Ser	Gly	Tyr	Lys	Glu	Asn	Tyr	Thr	Arg	Trp	Leu	Cys	Tyr
		275					280					285			
Cys	Asn	Val	Pro	Gln	Phe	Cys	Asp	Ser	Leu	Pro	Arg	Tyr	Glu	Thr	Thr
	290					295					300				
Gln	Val	Phe	Gly	Arg	Thr	Leu	Leu	Arg	Ser	Val	Phe	Thr	Val	Met	Arg
305					310					315				320	
Arg	Gln	Leu	Leu	Glu	Gln	Ala	Arg	Gln	Glu	Lys	Asp	Lys	Leu	Pro	Leu
				325					330					335	
Glu	Lys	Arg	Thr	Leu	Ile	Leu	Thr	His	Phe	Pro	Lys	Phe	Leu	Ser	
			340					345					350		

## (2) INFORMATION FOR SEQ ID NO:5:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 476 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: None

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Met	Leu	Glu	Glu	Glu	Ile	Tyr	Gly	Ala	Asn	Ser	Pro	Ile	Trp	Glu	Ser
1				5					10					15	
Gly	Phe	Thr	Met	Pro	Pro	Ser	Glu	Gly	Thr	Gln	Leu	Val	Pro	Arg	Pro
			20					25					30		
Ala	Ser	Val	Ser	Ala	Ala	Val	Val	Pro	Ser	Thr	Pro	Ile	Phe	Ser	Pro
		35					40					45			
Ser	Met	Gly	Gly	Gly	Ser	Asn	Ser	Ser	Leu	Ser	Leu	Asp	Ser	Ala	Gly
	50					55				60					
Ala	Glu	Pro	Met	Pro	Gly	Glu	Lys	Arg	Thr	Leu	Pro	Glu	Asn	Leu	Thr
65					70					75				80	
Leu	Glu	Asp	Ala	Lys	Arg	Leu	Arg	Val	Met	Gly	Asp	Ile	Pro	Met	Glu
				85					90					95	
Leu	Val	Asn	Glu	Val	Met	Leu	Thr	Ile	Thr	Asp	Pro	Ala	Ala	Met	Leu
			100					105					110		
Gly	Pro	Glu	Thr	Ser	Leu	Leu	Ser	Ala	Asn	Ala	Ala	Arg	Asp	Glu	Thr
		115					120					125			
Ala	Arg	Leu	Glu	Glu	Arg	Arg	Gly	Ile	Ile	Glu	Phe	His	Val	Ile	Gly
	130					135					140				
Asn	Ser	Leu	Thr	Pro	Lys	Ala	Asn	Arg	Arg	Val	Leu	Leu	Trp	Leu	Val
145					150					155					160
Gly	Leu	Gln	Asn	Val	Phe	Ser	His	Gln	Leu	Pro	Arg	Met	Pro	Lys	Glu
				165					170					175	
Tyr	Ile	Ala	Arg	Leu	Val	Phe	Asp	Pro	Lys	His	Lys	Thr	Leu	Ala	Leu
			180					185					190		

```

Ile Lys Asp Gly Arg Val Ile Gly Gly Ile Cys Phe Arg Met Phe Pro
195 200 205
Thr Gln Gly Phe Thr Glu Ile Val Phe Cys Ala Val Thr Ser Asn Glu
210 215 220
Gln Val Lys Gly Tyr Gly Thr His Leu Met Asn His Leu Lys Glu Tyr
225 230 235 240
His Ile Lys His Asn Ile Leu Tyr Phe Leu Thr Tyr Ala Asp Glu Tyr
245 250 255
Ala Ile Gly Tyr Phe Lys Lys Gln Gly Phe Ser Lys Asp Ile Lys Val
260 265 270
Pro Lys Ser Arg Tyr Leu Gly Tyr Ile Lys Asp Tyr Glu Gly Ala Thr
275 280 285
Leu Met Glu Cys Glu Leu Asn Pro Arg Ile Pro Tyr Thr Glu Leu Ser
290 295 300
His Ile Ile Lys Lys Gln Lys Glu Ile Ile Lys Lys Leu Ile Glu Arg
305 310 315 320
Lys Gln Ala Gln Ile Arg Lys Val Tyr Pro Gly Leu Ser Cys Phe Lys
325 330 335
Glu Gly Val Arg Gln Ile Pro Val Glu Ser Val Pro Gly Ile Arg Glu
340 345 350
Thr Gly Trp Lys Pro Leu Gly Lys Glu Lys Gly Lys Glu Lys Asp
355 360 365
Pro Asp Gln Leu Tyr Thr Thr Leu Lys Asn Leu Leu Ala Gln Ile Lys
370 375 380
Ser His Pro Ser Ala Trp Pro Phe Met Glu Pro Val Lys Lys Ser Glu
385 390 395 400
Ala Pro Asp Tyr Tyr Glu Val Ile Arg Phe Pro Ile Asp Leu Lys Thr
405 410 415
Met Thr Glu Arg Leu Arg Ser Arg Tyr Tyr Val Thr Arg Lys Leu Phe
420 425 430
Val Ala Asp Leu Gln Arg Val Ile Ala Asn Cys Arg Glu Tyr Asn Pro
435 440 445
Pro Asp Ser Glu Tyr Cys Arg Cys Ala Ser Ala Leu Glu Lys Phe Phe
450 455 460
Tyr Phe Lys Leu Lys Glu Gly Gly Leu Ile Asp Lys
465 470 475

```

## (2) INFORMATION FOR SEQ ID NO:6:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2414 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: None

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```

Met Ala Glu Asn Val Val Glu Pro Gly Pro Pro Ser Ala Lys Arg Pro
1 5 10 15
Lys Leu Ser Ser Pro Ala Leu Ser Ala Ser Asp Gly Thr Asp
20 25 30
Phe Gly Ser Leu Phe Asp Leu Glu His Asp Leu Pro Asp Glu Leu Ile
35 40 45
Asn Ser Thr Glu Leu Gly Leu Thr Asn Gly Gly Asp Ile Asn Gln Leu
50 55 60
Gln Thr Ser Leu Gly Met Val Gln Asp Ala Ala Ser Lys His Lys Gln
65 70 75 80
Leu Ser Glu Leu Leu Arg Ser Gly Ser Ser Pro Asn Leu Asn Met Gly
85 90 95
Val Gly Gly Pro Gly Gln Val Met Ala Ser Gln Ala Gln Gln Ser Ser
100 105 110

```

Pro	Gly	Leu	Gly	Leu	Ile	Asn	Ser	Met	Val	Lys	Ser	Pro	Met	Thr	Gln
		115					120					125			
Ala	Gly	Leu	Thr	Ser	Pro	Asn	Met	Gly	Met	Gly	Thr	Ser	Gly	Pro	Asn
	130					135					140				
Gln	Gly	Pro	Thr	Gln	Ser	Thr	Gly	Met	Met	Asn	Ser	Pro	Val	Asn	Gln
145					150					155					160
Pro	Ala	Met	Gly	Met	Asn	Thr	Gly	Thr	Asn	Ala	Gly	Met	Asn	Pro	Gly
			165						170					175	
Met	Leu	Ala	Ala	Gly	Asn	Gly	Gln	Gly	Ile	Met	Pro	Asn	Gln	Val	Met
		180						185					190		
Asn	Gly	Ser	Ile	Gly	Ala	Gly	Arg	Gly	Arg	Gln	Asp	Met	Gln	Tyr	Pro
	195						200					205			
Asn	Pro	Gly	Met	Gly	Ser	Ala	Gly	Asn	Leu	Leu	Thr	Glu	Pro	Leu	Gln
	210					215					220				
Gln	Gly	Ser	Pro	Gln	Met	Gly	Gly	Gln	Thr	Gly	Leu	Arg	Gly	Pro	Gln
225					230					235					240
Pro	Leu	Lys	Met	Gly	Met	Met	Asn	Asn	Pro	Asn	Pro	Tyr	Gly	Ser	Pro
			245						250					255	
Tyr	Thr	Gln	Asn	Pro	Gly	Gln	Gln	Ile	Gly	Ala	Ser	Gly	Leu	Gly	Leu
		260						265					270		
Gln	Ile	Gln	Thr	Lys	Thr	Val	Leu	Ser	Asn	Asn	Leu	Ser	Pro	Phe	Ala
	275						280					285			
Met	Asp	Lys	Lys	Ala	Val	Pro	Gly	Gly	Gly	Met	Pro	Asn	Met	Gly	Gln
	290					295					300				
Gln	Pro	Ala	Pro	Gln	Val	Gln	Gln	Pro	Gly	Leu	Val	Thr	Pro	Val	Ala
305					310					315					320
Gln	Gly	Met	Gly	Ser	Gly	Ala	His	Thr	Ala	Asp	Pro	Glu	Lys	Arg	Lys
			325						330					335	
Leu	Ile	Gln	Gln	Gln	Leu	Val	Leu	Leu	Leu	His	Ala	His	Lys	Cys	Gln
		340						345					350		
Arg	Arg	Glu	Gln	Ala	Asn	Gly	Glu	Val	Arg	Gln	Cys	Asn	Leu	Pro	His
		355					360					365			
Cys	Arg	Thr	Met	Lys	Asn	Val	Leu	Asn	His	Met	Thr	His	Cys	Gln	Ser
	370				375						380				
Gly	Lys	Ser	Cys	Gln	Val	Ala	His	Cys	Ala	Ser	Ser	Arg	Gln	Ile	Ile
385					390					395					400
Ser	His	Trp	Lys	Asn	Cys	Thr	Arg	His	Asp	Cys	Pro	Val	Cys	Leu	Pro
			405						410					415	
Leu	Lys	Asn	Ala	Gly	Asp	Lys	Arg	Asn	Gln	Gln	Pro	Ile	Leu	Thr	Gly
		420						425					430		
Ala	Pro	Val	Gly	Leu	Gly	Asn	Pro	Ser	Ser	Leu	Gly	Val	Gly	Gln	Gln
	435						440					445			
Ser	Ala	Pro	Asn	Leu	Ser	Thr	Val	Ser	Gln	Ile	Asp	Pro	Ser	Ser	Ile
	450					455					460				
Glu	Arg	Ala	Tyr	Ala	Ala	Leu	Gly	Leu	Pro	Tyr	Gln	Val	Asn	Gln	Met
465					470					475					480
Pro	Thr	Gln	Pro	Gln	Val	Gln	Ala	Lys	Asn	Gln	Gln	Asn	Gln	Gln	Pro
			485						490					495	
Gly	Gln	Ser	Pro	Gln	Gly	Met	Arg	Pro	Met	Ser	Asn	Met	Ser	Ala	Ser
		500						505					510		
Pro	Met	Gly	Val	Asn	Gly	Gly	Val	Gly	Val	Gln	Thr	Pro	Ser	Leu	Leu
	515						520					525			
Ser	Asp	Ser	Met	Leu	His	Ser	Ala	Ile	Asn	Ser	Gln	Asn	Pro	Met	Met
	530					535					540				
Ser	Glu	Asn	Ala	Ser	Val	Pro	Ser	Leu	Gly	Pro	Met	Pro	Thr	Ala	Ala
545					550					555					560
Gln	Pro	Ser	Thr	Thr	Gly	Ile	Arg	Lys	Gln	Trp	His	Glu	Asp	Ile	Thr
			565						570					575	
Gln	Asp	Leu	Arg	Asn	His	Leu	Val	His	Lys	Leu	Val	Gln	Ala	Ile	Phe
		580						585					590		
Pro	Thr	Pro	Asp	Pro	Ala	Ala	Leu	Lys	Asp	Arg	Arg	Met	Glu	Asn	Leu
		595					600					605			

Val	Ala	Tyr	Ala	Arg	Lys	Val	Glu	Gly	Asp	Met	Tyr	Glu	Ser	Ala	Asn
610						615					620				
Asn	Arg	Ala	Glu	Tyr	Tyr	His	Leu	Leu	Ala	Glu	Lys	Ile	Tyr	Lys	Ile
625					630					635					640
Gln	Lys	Glu	Leu	Glu	Glu	Lys	Arg	Arg	Thr	Arg	Leu	Gln	Lys	Gln	Asn
			645						650					655	
Met	Leu	Pro	Asn	Ala	Ala	Gly	Met	Val	Pro	Val	Ser	Met	Asn	Pro	Gly
			660					665					670		
Pro	Asn	Met	Gly	Gln	Pro	Gln	Pro	Gly	Met	Thr	Ser	Asn	Gly	Pro	Leu
		675					680					685			
Pro	Asp	Pro	Ser	Met	Ile	Arg	Gly	Ser	Val	Pro	Asn	Gln	Met	Met	Pro
	690					695					700				
Arg	Ile	Thr	Pro	Gln	Ser	Gly	Leu	Asn	Gln	Phe	Gly	Gln	Met	Ser	Met
705					710					715					720
Ala	Gln	Pro	Pro	Ile	Val	Pro	Arg	Gln	Thr	Pro	Pro	Leu	Gln	His	His
				725						730				735	
Gly	Gln	Leu	Ala	Gln	Pro	Gly	Ala	Leu	Asn	Pro	Pro	Met	Gly	Tyr	Gly
			740					745					750		
Pro	Arg	Met	Gln	Gln	Pro	Ser	Asn	Gln	Gly	Gln	Phe	Leu	Pro	Gln	Thr
		755					760					765			
Gln	Phe	Pro	Ser	Gln	Gly	Met	Asn	Val	Thr	Asn	Ile	Pro	Leu	Ala	Pro
	770					775					780				
Ser	Ser	Gly	Gln	Ala	Pro	Val	Ser	Gln	Ala	Gln	Met	Ser	Ser	Ser	Ser
785					790					795					800
Cys	Pro	Val	Asn	Ser	Pro	Ile	Met	Pro	Pro	Gly	Ser	Gln	Gly	Ser	His
			805						810					815	
Ile	His	Cys	Pro	Gln	Leu	Pro	Gln	Pro	Ala	Leu	His	Gln	Asn	Ser	Pro
			820					825					830		
Ser	Pro	Val	Pro	Ser	Arg	Thr	Pro	Thr	Pro	His	His	Thr	Pro	Pro	Ser
		835					840					845			
Ile	Gly	Ala	Gln	Gln	Pro	Pro	Ala	Thr	Thr	Ile	Pro	Ala	Pro	Val	Pro
	850					855					860				
Thr	Pro	Pro	Ala	Met	Pro	Pro	Gly	Pro	Gln	Ser	Gln	Ala	Leu	His	Pro
865					870					875					880
Pro	Pro	Arg	Gln	Thr	Pro	Thr	Pro	Pro	Thr	Thr	Gln	Leu	Pro	Gln	Gln
			885						890					895	
Val	Gln	Pro	Ser	Leu	Pro	Ala	Ala	Pro	Ser	Ala	Asp	Gln	Pro	Gln	Gln
			900					905					910		
Gln	Pro	Arg	Ser	Gln	Gln	Ser	Thr	Ala	Ala	Ser	Val	Pro	Thr	Pro	Asn
		915					920					925			
Ala	Pro	Leu	Leu	Pro	Pro	Gln	Pro	Ala	Thr	Pro	Leu	Ser	Gln	Pro	Ala
		930				935					940				
Val	Ser	Ile	Glu	Gly	Gln	Val	Ser	Asn	Pro	Pro	Ser	Thr	Ser	Ser	Thr
945					950					955					960
Glu	Val	Asn	Ser	Gln	Ala	Ile	Ala	Glu	Lys	Gln	Pro	Ser	Gln	Glu	Val
			965						970					975	
Lys	Met	Glu	Ala	Lys	Met	Glu	Val	Asp	Gln	Pro	Glu	Pro	Ala	Asp	Thr
			980					985					990		
Gln	Pro	Glu	Asp	Ile	Ser	Glu	Ser	Lys	Val	Glu	Asp	Cys	Lys	Met	Glu
		995				1000					1005				
Ser	Thr	Glu	Thr	Glu	Glu	Arg	Ser	Thr	Glu	Leu	Lys	Thr	Glu	Ile	Lys
	1010					1015					1020				
Glu	Glu	Glu	Asp	Gln	Pro	Ser	Thr	Ser	Ala	Thr	Gln	Ser	Ser	Pro	Ala
025				1030						1035					1040
Pro	Gly	Gln	Ser	Lys	Lys	Lys	Ile	Phe	Lys	Pro	Glu	Glu	Leu	Arg	Gln
			1045					1050						1055	
Ala	Leu	Met	Pro	Thr	Leu	Glu	Ala	Leu	Tyr	Arg	Gln	Asp	Pro	Glu	Ser
			1060					1065					1070		
Leu	Pro	Phe	Arg	Gln	Pro	Val	Asp	Pro	Gln	Leu	Leu	Gly	Ile	Pro	Asp
		1075					1080					1085			
Tyr	Phe	Asp	Ile	Val	Lys	Ser	Pro	Met	Asp	Leu	Ser	Thr	Ile	Lys	Arg
	1090					1095						1100			

Lys Leu Asp Thr Gly Gln Tyr Gln Glu Pro Trp Gln Tyr Val Asp Asp  
 105 1110 1115 1120  
 Ile Trp Leu Met Phe Asn Asn Ala Trp Leu Tyr Asn Arg Lys Thr Ser  
 1125 1130 1135  
 Arg Val Tyr Lys Tyr Cys Ser Lys Leu Ser Glu Val Phe Glu Gln Glu  
 1140 1145 1150  
 Ile Asp Pro Val Met Gln Ser Leu Gly Tyr Cys Cys Gly Arg Lys Leu  
 1155 1160 1165  
 Glu Phe Ser Pro Gln Thr Leu Cys Cys Tyr Gly Lys Gln Leu Cys Thr  
 1170 1175 1180  
 Ile Pro Arg Asp Ala Thr Tyr Tyr Ser Tyr Gln Asn Arg Tyr His Phe  
 185 1190 1195 1200  
 Cys Glu Lys Cys Phe Asn Glu Ile Gln Gly Glu Ser Val Ser Leu Gly  
 1205 1210 1215  
 Asp Asp Pro Ser Gln Pro Gln Thr Thr Ile Asn Lys Glu Gln Phe Ser  
 1220 1225 1230  
 Lys Arg Lys Asn Asp Thr Leu Asp Pro Glu Leu Phe Val Glu Cys Thr  
 1235 1240 1245  
 Glu Cys Gly Arg Lys Met His Gln Ile Cys Val Leu His His Glu Ile  
 1250 1255 1260  
 Ile Trp Pro Ala Gly Phe Val Cys Asp Gly Cys Leu Lys Lys Ser Ala  
 265 1270 1275 1280  
 Arg Thr Arg Lys Glu Asn Lys Phe Ser Ala Lys Arg Leu Pro Ser Thr  
 1285 1290 1295  
 Arg Leu Gly Thr Phe Leu Glu Asn Arg Val Asn Asp Phe Leu Arg Arg  
 1300 1305 1310  
 Gln Asn His Pro Glu Ser Gly Glu Val Thr Val Arg Val Val His Ala  
 1315 1320 1325  
 Ser Asp Lys Thr Val Glu Val Lys Pro Gly Met Lys Ala Arg Phe Val  
 1330 1335 1340  
 Asp Ser Gly Glu Met Ala Glu Ser Phe Pro Tyr Arg Thr Lys Ala Leu  
 345 1350 1355 1360  
 Phe Ala Phe Glu Glu Ile Asp Gly Val Asp Leu Cys Phe Phe Gly Met  
 1365 1370 1375  
 His Val Gln Glu Tyr Gly Ser Asp Cys Pro Pro Pro Asn Gln Arg Arg  
 1380 1385 1390  
 Val Tyr Ile Ser Tyr Leu Asp Ser Val His Phe Phe Arg Pro Lys Cys  
 1395 1400 1405  
 Leu Arg Thr Ala Val Tyr His Glu Ile Leu Ile Gly Tyr Leu Glu Tyr  
 1410 1415 1420  
 Val Lys Lys Leu Gly Tyr Thr Thr Gly His Ile Trp Ala Cys Pro Pro  
 425 1430 1435 1440  
 Ser Glu Gly Asp Asp Tyr Ile Phe His Cys His Pro Pro Asp Gln Lys  
 1445 1450 1455  
 Ile Pro Lys Pro Lys Arg Leu Gln Glu Trp Tyr Lys Lys Met Leu Asp  
 1460 1465 1470  
 Lys Ala Val Ser Glu Arg Ile Val His Asp Tyr Lys Asp Ile Phe Lys  
 1475 1480 1485  
 Gln Ala Thr Glu Asp Arg Leu Thr Ser Ala Lys Glu Leu Pro Tyr Phe  
 1490 1495 1500  
 Glu Gly Asp Phe Trp Pro Asn Val Leu Glu Glu Ser Ile Lys Glu Leu  
 505 1510 1515 1520  
 Glu Gln Glu Glu Glu Glu Arg Lys Arg Glu Glu Asn Thr Ser Asn Glu  
 1525 1530 1535  
 Ser Thr Asp Val Thr Lys Gly Asp Ser Lys Asn Ala Lys Lys Lys Asn  
 1540 1545 1550  
 Asn Lys Lys Thr Ser Lys Asn Lys Ser Ser Leu Ser Arg Gly Asn Lys  
 1555 1560 1565  
 Lys Lys Pro Gly Met Pro Asn Val Ser Asn Asp Leu Ser Gln Lys Leu  
 1570 1575 1580  
 Tyr Ala Thr Met Glu Lys His Lys Glu Val Phe Val Ile Arg Leu  
 585 1590 1595 1600

Ile Ala Gly Pro Ala Ala Asn Ser Leu Pro Pro Ile Val Asp Pro Asp  
 1605 1610 1615  
 Pro Leu Ile Pro Cys Asp Leu Met Asp Gly Arg Asp Ala Phe Leu Thr  
 1620 1625 1630  
 Leu Ala Arg Asp Lys His Leu Glu Phe Ser Ser Leu Arg Arg Ala Gln  
 1635 1640 1645  
 Trp Ser Thr Met Cys Met Leu Val Glu Leu His Thr Gln Ser Gln Asp  
 1650 1655 1660  
 Arg Phe Val Tyr Thr Cys Asn Glu Cys Lys His His Val Glu Thr Arg  
 665 1670 1675 1680  
 Trp His Cys Thr Val Cys Glu Asp Tyr Asp Leu Cys Ile Thr Cys Tyr  
 1685 1690 1695  
 Asn Thr Lys Asn His Asp His Lys Met Glu Lys Leu Gly Leu Gly Leu  
 1700 1705 1710  
 Asp Asp Glu Ser Asn Asn Gln Gln Ala Ala Thr Gln Ser Pro Gly  
 1715 1720 1725  
 Asp Ser Arg Arg Leu Ser Ile Gln Arg Cys Ile Gln Ser Leu Val His  
 1730 1735 1740  
 Ala Cys Gln Cys Arg Asn Ala Asn Cys Ser Leu Pro Ser Cys Gln Lys  
 745 1750 1755 1760  
 Met Lys Arg Val Val Gln His Thr Lys Gly Cys Lys Arg Lys Thr Asn  
 1765 1770 1775  
 Gly Gly Cys Pro Ile Cys Lys Gln Leu Ile Ala Leu Cys Cys Tyr His  
 1780 1785 1790  
 Ala Lys His Cys Gln Glu Asn Lys Cys Pro Val Pro Phe Cys Leu Asn  
 1795 1800 1805  
 Ile Lys Gln Lys Leu Arg Gln Gln Gln Leu Gln His Arg Leu Gln Gln  
 1810 1815 1820  
 Ala Gln Met Leu Arg Arg Arg Met Ala Ser Met Gln Arg Thr Gly Val  
 825 1830 1835 1840  
 Val Gly Gln Gln Gln Gly Leu Pro Ser Pro Thr Pro Ala Thr Pro Thr  
 1845 1850 1855  
 Thr Pro Thr Gly Gln Gln Pro Thr Thr Pro Gln Thr Pro Gln Pro Thr  
 1860 1865 1870  
 Ser Gln Pro Gln Pro Thr Pro Pro Asn Ser Met Pro Pro Tyr Leu Pro  
 1875 1880 1885  
 Arg Thr Gln Ala Ala Gly Pro Val Ser Gln Gly Lys Ala Ala Gly Gln  
 1890 1895 1900  
 Val Thr Pro Pro Thr Pro Gln Thr Ala Gln Pro Pro Leu Pro Gly  
 905 1910 1915 1920  
 Pro Pro Pro Thr Ala Val Glu Met Ala Met Gln Ile Gln Arg Ala Ala  
 1925 1930 1935  
 Glu Thr Gln Arg Gln Met Ala His Val Gln Ile Phe Gln Arg Pro Ile  
 1940 1945 1950  
 Gln His Gln Met Pro Pro Met Thr Pro Met Ala Pro Met Gly Met Asn  
 1955 1960 1965  
 Pro Pro Pro Met Thr Arg Gly Pro Ser Gly His Leu Glu Pro Gly Met  
 1970 1975 1980  
 Gly Pro Thr Gly Met Gln Gln Gln Pro Pro Trp Ser Gln Gly Gly Leu  
 985 1990 1995 2000  
 Pro Gln Pro Gln Gln Leu Gln Ser Gly Met Pro Arg Pro Ala Met Met  
 2005 2010 2015  
 Ser Val Ala Gln His Gly Gln Pro Leu Asn Met Ala Pro Gln Pro Gly  
 2020 2025 2030  
 Leu Gly Gln Val Gly Ile Ser Pro Leu Lys Pro Gly Thr Val Ser Gln  
 2035 2040 2045  
 Gln Ala Leu Gln Asn Leu Leu Arg Thr Leu Arg Ser Pro Ser Ser Pro  
 2050 2055 2060  
 Leu Gln Gln Gln Gln Val Leu Ser Ile Leu His Ala Asn Pro Gln Leu  
 065 2070 2075 2080  
 Leu Ala Ala Phe Ile Lys Gln Arg Ala Ala Lys Tyr Ala Asn Ser Asn  
 2085 2090 2095

Pro Gln Pro Ile Pro Gly Gln Pro Gly Met Pro Gln Gly Gln Pro Gly  
                   2100                  2105                  2110  
 Leu Gln Pro Pro Thr Met Pro Gly Gln Gln Gly Val His Ser Asn Pro  
                   2115                  2120                  2125  
 Ala Met Gln Asn Met Asn Pro Met Gln Ala Gly Val Gln Arg Ala Gly  
                   2130                  2135                  2140  
 Leu Pro Gln Gln Gln Pro Gln Gln Gln Leu Gln Pro Pro Met Gly Gly  
 145                  2150                  2155                  2160  
 Met Ser Pro Gln Ala Gln Gln Met Asn Met Asn His Asn Thr Met Pro  
                   2165                  2170                  2175  
 Ser Gln Phe Arg Asp Ile Leu Arg Arg Gln Gln Met Met Gln Gln Gln  
                   2180                  2185                  2190  
 Gln Gln Gln Gly Ala Gly Pro Gly Ile Gly Pro Gly Met Ala Asn His  
                   2195                  2200                  2205  
 Asn Gln Phe Gln Gln Pro Gln Gly Val Gly Tyr Pro Pro Gln Pro Gln  
                   2210                  2215                  2220  
 Gln Arg Met Gln His His Met Gln Gln Met Gln Gln Gly Asn Met Gly  
 225                  2230                  2235                  2240  
 Gln Ile Gly Gln Leu Pro Gln Ala Leu Gly Ala Glu Ala Gly Ala Ser  
                   2245                  2250                  2255  
 Leu Gln Ala Tyr Gln Gln Arg Leu Leu Gln Gln Gln Met Gly Ser Pro  
                   2260                  2265                  2270  
 Val Gln Pro Asn Pro Met Ser Pro Gln Gln His Met Leu Pro Asn Gln  
                   2275                  2280                  2285  
 Ala Gln Ser Pro His Leu Gln Gly Gln Gln Ile Pro Asn Ser Leu Ser  
                   2290                  2295                  2300  
 Asn Gln Val Arg Ser Pro Gln Pro Val Pro Ser Pro Arg Pro Gln Ser  
 305                  2310                  2315                  2320  
 Gln Pro Pro His Ser Ser Pro Ser Pro Arg Met Gln Pro Gln Pro Ser  
                   2325                  2330                  2335  
 Pro His His Val Ser Pro Gln Thr Ser Ser Pro His Pro Gly Leu Val  
                   2340                  2345                  2350  
 Ala Ala Gln Ala Asn Pro Met Glu Gln Gly His Phe Ala Ser Pro Asp  
                   2355                  2360                  2365  
 Gln Asn Ser Met Leu Ser Gln Leu Ala Ser Asn Pro Gly Met Ala Asn  
                   2370                  2375                  2380  
 Leu His Gly Ala Ser Ala Thr Asp Leu Gly Leu Ser Thr Asp Asn Ser  
 385                  2390                  2395                  2400  
 Asp Leu Asn Ser Asn Leu Ser Gln Ser Thr Leu Asp Ile His  
                   2405                  2410                  2

## (2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2441 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Met Ala Glu Asn Leu Leu Asp Gly Pro Pro Asn Pro Lys Arg Ala Lys  
 1                  5                  10                  15  
 Leu Ser Ser Pro Gly Phe Ser Ala Asn Asp Asn Thr Asp Phe Gly Ser  
                   20                  25                  30  
 Leu Phe Asp Leu Glu Asn Asp Leu Pro Asp Glu Leu Ile Pro Asn Gly  
                   35                  40                  45  
 Glu Leu Ser Leu Leu Asn Ser Gly Asn Leu Val Pro Asp Ala Ala Ser  
                   50                  55                  60  
 Lys His Lys Gln Leu Ser Glu Leu Leu Arg Gly Gly Ser Gly Ser Ser  
 65                  70                  75                  80

Ile	Asn	Pro	Gly	Ile	Gly	Asn	Val	Ser	Ala	Ser	Ser	Pro	Val	Gln	Gln
				85					90					95	
Gly	Leu	Gly	Gly	Gln	Ala	Gln	Gly	Gln	Pro	Asn	Ser	Thr	Asn	Met	Ala
			100					105					110		
Ser	Leu	Gly	Ala	Met	Gly	Lys	Ser	Pro	Leu	Asn	Gln	Gly	Asp	Ser	Ser
		115					120					125			
Thr	Pro	Asn	Leu	Pro	Lys	Gln	Ala	Ala	Ser	Thr	Ser	Gly	Pro	Thr	Pro
	130					135					140				
Pro	Ala	Ser	Gln	Ala	Leu	Asn	Pro	Gln	Ala	Gln	Lys	Gln	Val	Gly	Leu
145					150					155					160
Val	Thr	Ser	Ser	Pro	Ala	Thr	Ser	Gln	Thr	Gly	Pro	Gly	Ile	Cys	Met
				165					170					175	
Asn	Ala	Asn	Phe	Asn	Gln	Thr	His	Pro	Gly	Leu	Leu	Asn	Ser	Asn	Ser
			180					185					190		
Gly	His	Ser	Leu	Met	Asn	Gln	Ala	Gln	Gln	Gly	Gln	Ala	Gln	Val	Met
		195					200					205			
Asn	Gly	Ser	Leu	Gly	Ala	Ala	Gly	Arg	Gly	Arg	Gly	Ala	Gly	Met	Pro
	210					215					220				
Tyr	Pro	Ala	Pro	Ala	Met	Gln	Gly	Ala	Thr	Ser	Ser	Val	Leu	Ala	Glu
225					230					235					240
Thr	Leu	Thr	Gln	Val	Ser	Pro	Gln	Met	Ala	Gly	His	Ala	Gly	Leu	Asn
				245					250					255	
Thr	Ala	Gln	Ala	Gly	Gly	Met	Thr	Lys	Met	Gly	Met	Thr	Gly	Thr	Thr
			260					265					270		
Ser	Pro	Phe	Gly	Gln	Pro	Phe	Ser	Gln	Thr	Gly	Gly	Gln	Gln	Met	Gly
		275					280					285			
Ala	Thr	Gly	Val	Asn	Pro	Gln	Leu	Ala	Ser	Lys	Gln	Ser	Met	Val	Asn
	290				295						300				
Ser	Leu	Pro	Ala	Phe	Pro	Thr	Asp	Ile	Lys	Asn	Thr	Ser	Val	Thr	Thr
305					310					315					320
Val	Pro	Asn	Met	Ser	Gln	Leu	Gln	Thr	Ser	Val	Gly	Ile	Val	Pro	Thr
			325						330					335	
Gln	Ala	Ile	Ala	Thr	Gly	Pro	Thr	Ala	Asp	Pro	Glu	Lys	Arg	Lys	Leu
			340					345					350		
Ile	Gln	Gln	Gln	Leu	Val	Leu	Leu	His	Ala	His	Lys	Cys	Gln	Arg	
		355				360					365				
Arg	Glu	Gln	Ala	Asn	Gly	Glu	Val	Arg	Ala	Cys	Ser	Leu	Pro	His	Cys
	370				375						380				
Arg	Thr	Met	Lys	Asn	Val	Leu	Asn	His	Met	Thr	His	Cys	Gln	Ala	Pro
385				390						395					400
Lys	Ala	Cys	Gln	Val	Ala	His	Cys	Ala	Ser	Ser	Arg	Gln	Ile	Ile	Ser
			405						410					415	
His	Trp	Lys	Asn	Cys	Thr	Arg	His	Asp	Cys	Pro	Val	Cys	Leu	Pro	Leu
			420					425					430		
Lys	Asn	Ala	Ser	Asp	Lys	Arg	Asn	Gln	Gln	Thr	Ile	Leu	Gly	Ser	Pro
	435						440					445			
Ala	Ser	Gly	Ile	Gln	Asn	Thr	Ile	Gly	Ser	Val	Gly	Ala	Gly	Gln	Gln
	450					455					460				
Asn	Ala	Thr	Ser	Leu	Ser	Asn	Pro	Asn	Pro	Ile	Asp	Pro	Ser	Ser	Met
465					470					475					480
Gln	Arg	Ala	Tyr	Ala	Ala	Leu	Gly	Leu	Pro	Tyr	Met	Asn	Gln	Pro	Gln
			485						490					495	
Thr	Gln	Leu	Gln	Pro	Gln	Val	Pro	Gly	Gln	Gln	Pro	Ala	Gln	Pro	Pro
			500					505					510		
Ala	His	Gln	Gln	Met	Arg	Thr	Leu	Asn	Ala	Leu	Gly	Asn	Asn	Pro	Met
		515					520					525			
Ser	Val	Pro	Ala	Gly	Gly	Ile	Thr	Thr	Asp	Gln	Gln	Pro	Pro	Asn	Leu
	530					535						540			
Ile	Ser	Glu	Ser	Ala	Leu	Pro	Thr	Ser	Leu	Gly	Ala	Thr	Asn	Pro	Leu
545					550					555					560
Met	Asn	Asp	Gly	Ser	Asn	Ser	Gly	Asn	Ile	Gly	Ser	Leu	Ser	Thr	Ile
				565					570					575	

Pro	Thr	Ala	Ala	Pro	Pro	Ser	Ser	Thr	Gly	Val	Arg	Lys	Gly	Trp	His
			580					585					590		
Glu	His	Val	Thr	Gln	Asp	Leu	Arg	Ser	His	Leu	Val	His	Lys	Leu	Val
		595					600					605			
Gln	Ala	Ile	Phe	Pro	Thr	Pro	Asp	Pro	Ala	Ala	Leu	Lys	Asp	Arg	Arg
	610					615					620				
Met	Glu	Asn	Leu	Val	Ala	Tyr	Ala	Lys	Lys	Val	Glu	Gly	Asp	Met	Tyr
625					630					635					640
Glu	Ser	Ala	Asn	Ser	Arg	Asp	Glu	Tyr	Tyr	His	Leu	Leu	Ala	Glu	Lys
			645						650					655	
Ile	Tyr	Lys	Ile	Gln	Lys	Glu	Leu	Glu	Glu	Lys	Arg	Arg	Thr	Arg	Leu
		660						665					670		
His	Lys	Gln	Gly	Ile	Leu	Gly	Asn	Gln	Pro	Ala	Leu	Pro	Ala	Ser	Gly
		675					680					685			
Ala	Gln	Pro	Pro	Val	Ile	Pro	Pro	Ala	Gln	Ser	Val	Arg	Pro	Pro	Asn
	690					695					700				
Gly	Pro	Leu	Pro	Leu	Pro	Val	Asn	Arg	Met	Gln	Val	Ser	Gln	Gly	Met
705					710					715					720
Asn	Ser	Phe	Asn	Pro	Met	Ser	Leu	Gly	Asn	Val	Gln	Leu	Pro	Gln	Ala
			725						730					735	
Pro	Met	Gly	Pro	Arg	Ala	Ala	Ser	Pro	Met	Asn	His	Ser	Val	Gln	Met
			740					745					750		
Asn	Ser	Met	Ala	Ser	Val	Pro	Gly	Met	Ala	Ile	Ser	Pro	Ser	Arg	Met
		755					760					765			
Pro	Gln	Pro	Pro	Asn	Met	Met	Gly	Thr	His	Ala	Asn	Asn	Ile	Met	Ala
	770					775					780				
Gln	Ala	Pro	Thr	Gln	Asn	Gln	Phe	Leu	Pro	Gln	Asn	Gln	Phe	Pro	Ser
785					790					795					800
Ser	Ser	Gly	Ala	Met	Ser	Val	Asn	Ser	Val	Gly	Met	Gly	Gln	Pro	Ala
			805						810					815	
Ala	Gln	Ala	Gly	Val	Ser	Gln	Gly	Gln	Glu	Pro	Gly	Ala	Ala	Leu	Pro
			820					825					830		
Asn	Pro	Leu	Asn	Met	Leu	Ala	Pro	Gln	Ala	Ser	Gln	Leu	Pro	Cys	Pro
		835					840					845			
Pro	Val	Thr	Gln	Ser	Pro	Leu	His	Pro	Thr	Pro	Pro	Pro	Ala	Ser	Thr
	850					855					860				
Ala	Ala	Gly	Met	Pro	Ser	Leu	Gln	His	Pro	Thr	Ala	Pro	Gly	Met	Thr
865					870					875					880
Pro	Pro	Gln	Pro	Ala	Ala	Pro	Thr	Gln	Pro	Ser	Thr	Pro	Val	Ser	Ser
			885						890					895	
Gly	Gln	Thr	Pro	Thr	Pro	Thr	Pro	Gly	Ser	Val	Pro	Ser	Ala	Ala	Gln
		900						905					910		
Thr	Gln	Ser	Thr	Pro	Thr	Val	Gln	Ala	Ala	Ala	Gln	Ala	Gln	Val	Thr
	915						920					925			
Pro	Gln	Pro	Gln	Thr	Pro	Val	Gln	Pro	Pro	Ser	Val	Ala	Thr	Pro	Gln
	930					935					940				
Ser	Ser	Gln	Gln	Gln	Pro	Thr	Pro	Val	His	Thr	Gln	Pro	Pro	Gly	Thr
945					950					955					960
Pro	Leu	Ser	Gln	Ala	Ala	Ser	Ile	Asp	Asn	Arg	Val	Pro	Thr	Pro	
			965					970					975		
Ser	Thr	Val	Thr	Ser	Ala	Glu	Thr	Ser	Ser	Gln	Gln	Pro	Gly	Pro	Asp
			980					985					990		
Val	Pro	Met	Leu	Glu	Met	Lys	Thr	Glu	Val	Gln	Thr	Asp	Asp	Ala	Glu
		995				1000						1005			
Pro	Glu	Pro	Thr	Glu	Ser	Lys	Gly	Glu	Pro	Arg	Ser	Glu	Met	Met	Glu
	1010					1015					1020				
Glu	Asp	Leu	Gln	Gly	Ser	Ser	Gln	Val	Lys	Glu	Glu	Thr	Asp	Thr	Thr
025					1030					1035					1040
Glu	Gln	Lys	Ser	Glu	Pro	Met	Glu	Val	Glu	Glu	Lys	Lys	Pro	Glu	Val
			1045						1050					1055	
Lys	Val	Glu	Ala	Lys	Glu	Glu	Glu	Glu	Asn	Ser	Ser	Asn	Asp	Thr	Ala
			1060					1065					1070		

Ser Gln Ser Thr Ser Pro Ser Gln Pro Arg Lys Lys Ile Phe Lys Pro  
 1075 1080 1085  
 Glu Glu Leu Arg Gln Ala Leu Met Pro Thr Leu Glu Ala Leu Tyr Arg  
 1090 1095 1100  
 Gln Asp Pro Glu Ser Leu Pro Phe Arg Gln Pro Val Asp Pro Gln Leu  
 105 1110 1115 1120  
 Leu Gly Ile Pro Asp Tyr Phe Asp Ile Val Lys Asn Pro Met Asp Leu  
 1125 1130 1135  
 Ser Thr Ile Lys Arg Lys Leu Asp Thr Gly Gln Tyr Gln Glu Pro Trp  
 1140 1145 1150  
 Gln Tyr Val Asp Asp Val Arg Leu Met Phe Asn Asn Ala Trp Leu Tyr  
 1155 1160 1165  
 Asn Arg Lys Thr Ser Arg Val Tyr Lys Phe Cys Ser Lys Leu Ala Glu  
 1170 1175 1180  
 Val Phe Glu Gln Glu Ile Asp Pro Val Met Gln Ser Leu Gly Tyr Cys  
 185 1190 1195 1200  
 Cys Gly Arg Lys Tyr Glu Phe Ser Pro Gln Thr Leu Cys Cys Tyr Gly  
 1205 1210 1215  
 Lys Gln Leu Cys Thr Ile Pro Arg Asp Ala Ala Tyr Tyr Ser Tyr Gln  
 1220 1225 1230  
 Asn Arg Tyr His Phe Cys Gly Lys Cys Phe Thr Glu Ile Gln Gly Glu  
 1235 1240 1245  
 Asn Val Thr Leu Gly Asp Asp Pro Ser Gln Pro Gln Thr Thr Ile Ser  
 1250 1255 1260  
 Lys Asp Gln Phe Glu Lys Lys Lys Asn Asp Thr Leu Asp Pro Glu Pro  
 265 1270 1275 1280  
 Phe Val Asp Cys Lys Glu Cys Gly Arg Lys Met His Gln Ile Cys Val  
 1285 1290 1295  
 Leu His Tyr Asp Ile Ile Trp Pro Ser Gly Phe Val Cys Asp Asn Cys  
 1300 1305 1310  
 Leu Lys Lys Thr Gly Arg Pro Arg Lys Glu Asn Lys Phe Ser Ala Lys  
 1315 1320 1325  
 Arg Leu Gln Thr Thr Arg Leu Gly Asn His Leu Glu Asp Arg Val Asn  
 1330 1335 1340  
 Lys Phe Leu Arg Arg Gln Asn His Pro Glu Ala Gly Glu Val Phe Val  
 345 1350 1355 1360  
 Arg Val Val Ala Ser Ser Asp Lys Thr Val Glu Val Lys Pro Gly Met  
 1365 1370 1375  
 Lys Ser Arg Phe Val Asp Ser Gly Glu Met Ser Glu Ser Phe Pro Tyr  
 1380 1385 1390  
 Arg Thr Lys Ala Leu Phe Ala Phe Glu Glu Ile Asp Gly Val Asp Val  
 1395 1400 1405  
 Cys Phe Phe Gly Met His Val Gln Asp Thr Ala Leu Ile Ala Pro His  
 1410 1415 1420  
 Gln Ile Gln Gly Cys Val Tyr Ile Ser Tyr Leu Asp Ser Ile His Phe  
 425 1430 1435 1440  
 Phe Arg Pro Arg Cys Leu Arg Thr Ala Val Tyr His Glu Ile Leu Ile  
 1445 1450 1455  
 Gly Tyr Leu Glu Tyr Val Lys Lys Leu Val Tyr Val Thr Ala His Ile  
 1460 1465 1470  
 Trp Ala Cys Pro Pro Ser Glu Gly Asp Asp Tyr Ile Phe His Cys His  
 1475 1480 1485  
 Pro Pro Asp Gln Lys Ile Pro Lys Pro Lys Arg Leu Gln Glu Trp Tyr  
 1490 1495 1500  
 Lys Lys Met Leu Asp Lys Ala Phe Ala Glu Arg Ile Ile Asn Asp Tyr  
 505 1510 1515 1520  
 Lys Asp Ile Phe Lys Gln Ala Asn Glu Asp Arg Leu Thr Ser Ala Lys  
 1525 1530 1535  
 Glu Leu Pro Tyr Phe Glu Gly Asp Phe Trp Pro Asn Val Leu Glu Glu  
 1540 1545 1550  
 Ser Ile Lys Glu Leu Glu Gln Glu Glu Glu Arg Lys Lys Glu Glu  
 1555 1560 1565

Ser Thr Ala Ala Ser Glu Thr Pro Glu Gly Ser Gln Gly Asp Ser Lys  
 1570 1575 1580  
 Asn Ala Lys Lys Lys Asn Asn Lys Lys Thr Asn Lys Asn Lys Ser Ser  
 585 1590 1595 1600  
 Ile Ser Arg Ala Asn Lys Lys Lys Pro Ser Met Pro Asn Val Ser Asn  
 1605 1610 1615  
 Asp Leu Ser Gln Lys Leu Tyr Ala Thr Met Glu Lys His Lys Glu Val  
 1620 1625 1630  
 Phe Phe Val Ile His Leu His Ala Gly Pro Val Ile Ser Thr Gln Pro  
 1635 1640 1645  
 Pro Ile Val Asp Pro Asp Pro Leu Leu Ser Cys Asp Leu Met Asp Gly  
 1650 1655 1660  
 Arg Asp Ala Phe Leu Thr Leu Ala Arg Asp Lys His Trp Glu Phe Ser  
 665 1670 1675 1680  
 Ser Leu Arg Arg Ser Lys Trp Ser Thr Leu Cys Met Leu Val Glu Leu  
 1685 1690 1695  
 His Thr Gln Gly Gln Asp Arg Phe Val Tyr Thr Cys Asn Glu Cys Lys  
 1700 1705 1710  
 His His Val Glu Thr Arg Trp His Cys Thr Val Cys Glu Asp Tyr Asp  
 1715 1720 1725  
 Leu Cys Ile Asn Cys Tyr Asn Thr Lys Ser His Thr His Lys Met Val  
 1730 1735 1740  
 Lys Trp Gly Leu Gly Leu Asp Asp Glu Gly Ser Ser Gln Gly Glu Pro  
 745 1750 1755 1760  
 Gln Ser Lys Ser Pro Gln Glu Ser Arg Arg Leu Ser Ile Gln Arg Cys  
 1765 1770 1775  
 Ile Gln Ser Leu Val His Ala Cys Gln Cys Arg Asn Ala Asn Cys Ser  
 1780 1785 1790  
 Leu Pro Ser Cys Gln Lys Met Lys Arg Val Val Gln His Thr Lys Gly  
 1795 1800 1805  
 Cys Lys Arg Lys Thr Asn Gly Gly Cys Pro Val Cys Lys Gln Leu Ile  
 1810 1815 1820  
 Ala Leu Cys Cys Tyr His Ala Lys His Cys Gln Glu Asn Lys Cys Pro  
 825 1830 1835 1840  
 Val Pro Phe Cys Leu Asn Ile Lys His Asn Val Arg Gln Gln Gln Ile  
 1845 1850 1855  
 Gln His Cys Leu Gln Gln Ala Gln Leu Met Arg Arg Arg Met Ala Thr  
 1860 1865 1870  
 Met Asn Thr Arg Asn Val Pro Gln Gln Ser Leu Pro Ser Pro Thr Ser  
 1875 1880 1885  
 Ala Pro Pro Gly Thr Pro Thr Gln Gln Pro Ser Thr Pro Gln Thr Pro  
 1890 1895 1900  
 Gln Pro Pro Ala Gln Pro Gln Pro Ser Pro Val Asn Met Ser Pro Ala  
 905 1910 1915 1920  
 Gly Phe Pro Asn Val Ala Arg Thr Gln Pro Pro Thr Ile Val Ser Ala  
 1925 1930 1935  
 Gly Lys Pro Thr Asn Gln Val Pro Ala Pro Pro Pro Ala Gln Pro  
 1940 1945 1950  
 Pro Pro Ala Ala Val Glu Ala Ala Arg Gln Ile Glu Arg Glu Ala Gln  
 1955 1960 1965  
 Gln Gln Gln His Leu Tyr Arg Ala Asn Ile Asn Asn Gly Met Pro Pro  
 1970 1975 1980  
 Gly Arg Asp Gly Met Gly Thr Pro Gly Ser Gln Met Thr Pro Val Gly  
 985 1990 1995 2000  
 Leu Asn Val Pro Arg Pro Asn Gln Val Ser Gly Pro Val Met Ser Ser  
 2005 2010 2015  
 Met Pro Pro Gly Gln Trp Gln Gln Ala Pro Ile Pro Gln Gln Gln Pro  
 2020 2025 2030  
 Met Pro Gly Met Pro Arg Pro Val Met Ser Met Gln Ala Gln Ala Ala  
 2035 2040 2045  
 Val Ala Gly Pro Arg Met Pro Asn Val Gln Pro Asn Arg Ser Ile Ser  
 2050 2055 2060

Pro Ser Ala Leu Gln Asp Leu Leu Arg Thr Leu Lys Ser Pro Ser Ser  
 065 2070 2075 2080  
 Pro Gln Gln Gln Gln Val Leu Asn Ile Leu Lys Ser Asn Pro Gln  
 2085 2090 2095  
 Leu Met Ala Ala Phe Ile Lys Gln Arg Thr Ala Lys Tyr Val Ala Asn  
 2100 2105 2110  
 Gln Pro Gly Met Gln Pro Gln Pro Gly Leu Gln Ser Gln Pro Gly Met  
 2115 2120 2125  
 Gln Pro Gln Pro Gly Met His Gln Gln Pro Ser Leu Gln Asn Leu Asn  
 2130 2135 2140  
 Ala Met Gln Ala Gly Val Pro Arg Pro Gly Val Pro Pro Pro Gln Pro  
 145 2150 2155 2160  
 Ala Met Gly Gly Leu Asn Pro Gln Gly Gln Ala Leu Asn Ile Met Asn  
 2165 2170 2175  
 Pro Gly His Asn Pro Asn Met Thr Asn Met Asn Pro Gln Tyr Arg Glu  
 2180 2185 2190  
 Met Val Arg Arg Gln Leu Leu Gln His Gln Gln Gln Gln Gln Gln  
 2195 2200 2205  
 Gln Gln Gln Gln Gln Gln Gln Gln Asn Ser Ala Ser Leu Ala Gly Gly  
 2210 2215 2220  
 Met Ala Gly His Ser Gln Phe Gln Gln Pro Gln Gly Pro Gly Gly Tyr  
 225 2230 2235 2240  
 Ala Pro Ala Met Gln Gln Gln Arg Met Gln Gln His Leu Pro Ile Gln  
 2245 2250 2255  
 Gly Ser Ser Met Gly Gln Met Ala Ala Pro Met Gly Gln Leu Gly Gln  
 2260 2265 2270  
 Met Gly Gln Pro Gly Leu Gly Ala Asp Ser Thr Pro Asn Ile Gln Gln  
 2275 2280 2285  
 Ala Leu Gln Gln Arg Ile Leu Gln Gln Gln Gln Met Lys Gln Gln Ile  
 2290 2295 2300  
 Gly Ser Pro Gly Gln Pro Asn Pro Met Ser Pro Gln Gln His Met Leu  
 305 2310 2315 2320  
 Ser Gly Gln Pro Gln Ala Ser His Leu Pro Gly Gln Gln Ile Ala Thr  
 2325 2330 2335  
 Ser Leu Ser Asn Gln Val Arg Ser Pro Ala Pro Val Gln Ser Pro Arg  
 2340 2345 2350  
 Pro Gln Ser Gln Pro Pro His Ser Ser Pro Ser Pro Arg Ile Gln Pro  
 2355 2360 2365  
 Gln Pro Ser Pro His His Val Ser Pro Gln Thr Gly Thr Pro His Pro  
 2370 2375 2380  
 Gly Leu Ala Val Thr Met Ala Ser Ser Met Asp Gln Gly His Leu Gly  
 385 2390 2395 2400  
 Asn Pro Glu Gln Ser Ala Met Leu Pro Gln Leu Asn Thr Pro Asn Arg  
 2405 2410 2415  
 Ser Ala Leu Ser Ser Glu Leu Ser Leu Val Gly Asp Thr Thr Gly Asp  
 2420 2425 2430  
 Thr Leu Glu Lys Phe Val Glu Gly Leu  
 2435 2440

## (2) INFORMATION FOR SEQ ID NO:8:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 813 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: None

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Met Ala Glu Ala Gly Gly Ala Gly Ser Pro Ala Leu Pro Pro Ala Pro  
 1 5 10 15

Pro	His	Gly	Ser	Pro	Arg	Thr	Leu	Ala	Thr	Ala	Ala	Gly	Ser	Ser	Ala		
			20					25					30				
Ser	Cys	Gly	Pro	Ala	Thr	Pro	Val	Ala	Ala	Ala	Gly	Thr	Ala	Glu	Gly		
		35					40					45					
Pro	Gly	Gly	Gly	Gly	Ser	Ala	Arg	Ile	Ala	Val	Lys	Lys	Ala	Gln	Leu		
	50				55						60						
Arg	Ser	Ala	Pro	Arg	Ala	Lys	Lys	Leu	Glu	Lys	Leu	Gly	Val	Tyr	Ser		
65					70					75				80			
Ala	Cys	Lys	Ala	Glu	Glu	Ser	Cys	Lys	Cys	Asn	Gly	Trp	Lys	Asn	Pro		
			85						90					95			
Asn	Pro	Ser	Pro	Thr	Pro	Pro	Arg	Gly	Asp	Leu	Gln	Gln	Ile	Ile	Val		
			100					105					110				
Ser	Leu	Thr	Glu	Ser	Cys	Arg	Ser	Cys	Ser	His	Ala	Leu	Ala	Ala	His		
		115					120					125					
Val	Ser	His	Leu	Glu	Asn	Val	Ser	Glu	Glu	Glu	Met	Asp	Arg	Leu	Leu		
	130				135						140						
Gly	Ile	Val	Leu	Asp	Val	Glu	Tyr	Leu	Phe	Thr	Cys	Val	His	Lys	Glu		
145				150						155					160		
Glu	Asp	Ala	Asp	Thr	Lys	Gln	Val	Tyr	Phe	Tyr	Leu	Phe	Lys	Leu	Leu		
			165						170					175			
Arg	Lys	Ser	Ile	Leu	Gln	Arg	Gly	Lys	Pro	Val	Val	Glu	Gly	Ser	Leu		
			180					185						190			
Glu	Lys	Lys	Pro	Pro	Phe	Glu	Lys	Pro	Ser	Ile	Glu	Gln	Gly	Val	Asn		
		195					200						205				
Asn	Phe	Val	Gln	Tyr	Lys	Phe	Ser	His	Leu	Pro	Ser	Lys	Glu	Arg	Gln		
	210					215						220					
Thr	Thr	Ile	Glu	Leu	Ala	Lys	Met	Phe	Leu	Asn	Arg	Ile	Asn	Tyr	Trp		
225				230						235					240		
His	Leu	Glu	Ala	Pro	Ser	Gln	Arg	Arg	Leu	Arg	Ser	Pro	Asn	Asp	Asp		
			245					250						255			
Ile	Ser	Gly	Tyr	Lys	Glu	Asn	Tyr	Thr	Arg	Trp	Leu	Cys	Tyr	Cys	Asn		
		260						265					270				
Val	Pro	Gln	Phe	Cys	Asp	Ser	Leu	Pro	Arg	Tyr	Glu	Thr	Thr	Lys	Val		
		275					280					285					
Phe	Gly	Arg	Thr	Leu	Leu	Arg	Ser	Val	Phe	Thr	Ile	Met	Arg	Arg	Gln		
	290					295					300						
Leu	Leu	Glu	Gln	Ala	Arg	Gln	Lys	Lys	Asp	Lys	Leu	Pro	Leu	Glu	Lys		
305				310						315					320		
Arg	Thr	Leu	Ile	Leu	Thr	His	Phe	Pro	Lys	Phe	Leu	Ser	Met	Leu	Glu		
			325						330					335			
Glu	Glu	Val	Tyr	Ser	Gln	Asn	Ser	Pro	Ile	Trp	Asp	Gln	Asp	Phe	Leu		
		340						345					350				
Ser	Ala	Ser	Ser	Arg	Thr	Ser	Pro	Leu	Gly	Ile	Gln	Thr	Val	Ile	Ser		
		355					360					365					
Pro	Pro	Val	Thr	Gly	Thr	Ala	Leu	Phe	Ser	Ser	Asn	Ser	Thr	Ser	His		
		370				375						380					
Glu	Gln	Ile	Asn	Gly	Gly	Arg	Thr	Ser	Pro	Gly	Cys	Arg	Gly	Ser	Ser		
385				390						395					400		
Gly	Leu	Glu	Ala	Asn	Pro	Gly	Glu	Lys	Arg	Lys	Met	Asn	Asn	Ser	His		
			405						410					415			
Ala	Pro	Glu	Glu	Ala	Lys	Arg	Ser	Arg	Val	Met	Gly	Asp	Ile	Pro	Val		
			420					425					430				
Glu	Leu	Ile	Asn	Glu	Val	Met	Ser	Thr	Ile	Thr	Asp	Pro	Ala	Gly	Met		
		435				440						445					
Leu	Gly	Pro	Glu	Thr	Asn	Phe	Leu	Ser	Ala	His	Ser	Ala	Arg	Asp	Glu		
	450					455					460						
Ala	Ala	Arg	Leu	Glu	Glu	Arg	Arg	Gly	Val	Ile	Glu	Phe	His	Val	Val		
465				470						475					480		
Gly	Asn	Ser	Leu	Asn	Gln	Lys	Pro	Asn	Lys	Lys	Ile	Leu	Met	Trp	Leu		
			485					490						495			
Val	Gly	Leu	Gln	Asn	Val	Phe	Ser	His	Gln	Leu	Pro	Arg	Met	Pro	Lys		
			500					505						510			

88

Glu Tyr Ile Thr Arg Leu Val Phe Asp Pro Lys His Lys Thr Leu Ala  
 515 520 525  
 Leu Ile Lys Asp Gly Arg Val Ile Gly Gly Ile Cys Phe Arg Met Phe  
 530 535 540  
 Pro Ser Gln Gly Phe Thr Glu Ile Val Phe Cys Ala Val Thr Ser Asn  
 545 550 555 560  
 Glu Gln Val Lys Gly Tyr Gly Thr His Leu Met Asn His Leu Lys Glu  
 565 570 575  
 Tyr His Ile Lys His Glu Ile Leu Asn Phe Leu Thr Tyr Ala Asp Glu  
 580 585 590  
 Tyr Ala Ile Gly Tyr Phe Lys Lys Gln Gly Phe Ser Lys Glu Ile Lys  
 595 600 605  
 Ile Pro Lys Thr Lys Tyr Val Gly Tyr Ile Lys Asp Tyr Glu Gly Ala  
 610 615 620  
 Thr Leu Met Gly Cys Glu Leu Asn Pro Gln Ile Pro Tyr Thr Glu Phe  
 625 630 635 640  
 Ser Val Ile Ile Lys Lys Gln Lys Glu Ile Ile Lys Lys Leu Ile Glu  
 645 650 655  
 Arg Lys Gln Ala Gln Ile Arg Lys Val Tyr Pro Gly Leu Ser Cys Phe  
 660 665 670  
 Lys Asp Gly Val Arg Gln Ile Pro Ile Glu Ser Ile Pro Gly Ile Arg  
 675 680 685  
 Glu Thr Gly Trp Lys Pro Ser Gly Lys Glu Lys Ser Lys Glu Pro Lys  
 690 695 700  
 Asp Pro Glu His Val Tyr Ser Thr Leu Lys Asn Ile Leu Gln Gln Val  
 705 710 715 720  
 Lys Asn His Pro Asn Ala Trp Pro Phe Met Glu Pro Val Lys Arg Thr  
 725 730 735  
 Glu Ala Pro Gly Tyr Tyr Glu Val Ile Arg Phe Pro Met Asp Leu Lys  
 740 745 750  
 Thr Met Ser Glu Arg Leu Arg Asn Arg Tyr Tyr Val Ser Lys Lys Leu  
 755 760 765  
 Phe Met Ala Asp Leu Gln Arg Val Phe Thr Asn Cys Lys Glu Tyr Asn  
 770 775 780  
 Pro Pro Glu Ser Glu Tyr Tyr Lys Cys Ala Ser Ile Leu Glu Lys Phe  
 785 790 795 800  
 Phe Phe Ser Lys Ile Lys Glu Ala Gly Leu Ile Asp Lys  
 805 810

## (2) INFORMATION FOR SEQ ID NO:9:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

His Thr Lys Gly Cys Lys Arg Lys Thr Asn Gly Gly Cys Pro Val Cys  
 1 5 10 15  
 Lys Gln Leu Ile Ala Leu Cys Cys Tyr His Ala Lys His Cys Gln Glu  
 20 25 30  
 Asn Lys Cys Pro Val Pro Phe Cys Leu Asn Ile Lys His Asn Val Arg  
 35 40 45  
 Gln Gln  
 50

## (2) INFORMATION FOR SEQ ID NO:10:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2204 base pairs

- (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

ACCCACTCCC	CCCAGAGCCG	ACCTGCAGCA	AATAATTGTC	AGTCTAACAG	AATCCTGTCTG	60
GAGTTGTAGC	CATGCCCTAG	CTGCTCATGT	TTCCCACCTG	GAGAATGTGT	CAGAGGAAGA	120
AATGAACAGA	CTCCTGGGAA	TAGTATTGGA	TGTGGAATAT	CTCTTTACCT	GTGTCCACAA	180
GGAAGAAGAT	GCAGATACCA	AACAAGTTTA	TTTCTATCTA	TTTAAGCTCT	TGAGAAAAGTC	240
TATTTTACAA	AGAGGAAAAC	CTGTGGTTGG	AAGGCTCTTT	GGAAAAGAAA	CCCCCATTTG	300
AAAAACCTAG	CATTGAACAG	GGTGTGAATA	ACTTTGTGCA	GTACAAATTT	AGTCACCTGC	360
CAGCAAAAAG	AAAGGCCAAAC	CAATAGTTGA	GTTGGCAAAA	ATGTTCCCTAA	ACCGCATCAC	420
CTATTGGCAT	CTGGAGGCAC	CATCTCAACG	AGACTGCGAT	CTCCAATGAT	GATATTCTGG	480
ATACAAAGAG	AACTACACAA	GGTGGCTGTG	TTACTGCAAC	GTGCCACAGT	TCTGCGACAG	540
TCTACCTCGG	TACGAAACCA	CACAGGTGTT	TGGGAGAACA	TCGTTCGCTC	GGTCTTCACT	600
GTTATGAGGC	GACAACTCCT	GGAACAAGCA	AGACAGGAAA	AAGATAAACT	GCCTCTTGAA	660
AAACGAACCT	TAATCCTCAC	TCATTTCCCA	AAATTTCTGT	CCATGCTAGA	AGAAGAAGTA	720
TATAGTCAAA	ACTCTCCCAT	CTGGGATCAC	CATTTTCTCT	CAGCCTCTTC	CAGAACCAGC	780
CAGCTAGGCA	TCCAAACAGT	TATCAATCAC	CTCCTGTGGC	TGGGACAATT	TCATACAATT	840
CAACCTCATC	TTCCCTTGAG	CAGCCAAACG	CAGGGAGCAG	CAGTCCTGCC	TGCAAAGCCT	900
CTTCTGGACT	TGAGGCCAAAC	CCAGGAGAAA	AGAGGAAAAT	GACTGATTCT	CATGTTCTGG	960
AGGAGGCCAA	GAAACCCCGA	GTTATGGGGG	ATATTCCGAT	GGAATTAATC	AACGAGGTTA	1020
TGTCTACCAT	CACGGACCCCT	GCAGCAATGC	TTGGACCAGA	GACCAATTTT	CTGTCAGCAC	1080
ACTCGGCCAG	GGATGAGGCG	GCAAGGTTGG	AAGAGCGCAG	GGGTGTAATT	GAATTTACAG	1140
TGGTTGGCAA	TTCCCTCAAC	CAGAAACCAA	ACAAGAAGAT	CCTGATGTGG	CTGGTTGGCC	1200
TACAGAACGT	TTTCTCCAC	CAGCTGCCCC	GAATGCCAAA	AGAATACATC	ACACGGCTCG	1260
TCTTTGACCC	GAAACACAAA	ACCCTTGCTT	TAATTAAAGA	TGGCCGTGTT	ATTGGTGGTA	1320
TCTGTTTCCG	TATGTTCCCA	TCTCAAGGAT	TCACAGAGAT	TGTCTTCTGT	GCTGTAACCT	1380
CAAATGAGCA	AGTCAAGGGC	TATGGAACAC	ACCTGATGAA	TCATTTGAAA	GAATATCACA	1440
TAAAGCATGA	CATCTGAAC	TTCTTACAT	ATGCAGATGA	ATATGCAATT	GGTACTTTTA	1500
AGAAACAGGG	TTTCTCCAAA	GAAATTAATA	TACCTAAAAC	CAAATATGTT	GGCTATATCA	1560
AGGATTATGA	AGGAGCCACT	TTAATGGGAT	GTGAGCTAAA	TCCACGGATC	CCGTACACAG	1620
AATTTTCTGT	CATCATTAATA	AAGCAGAAGG	AGATAATTAA	AAAACCTGAT	GAAAGAAAAC	1680
AGGCACAAAT	TCGAAAAGTT	TACCCTGGAC	TTTCATGTTT	TAAAGATGGA	GTTCGACAGA	1740
TTCTATAGA	AAGCATTCCT	GGAATTAGAG	AGACAGGCTG	GAAACCGAGT	GGAAAAGAGA	1800
AAAGTAAAGA	GCCCAGAGAC	CCTGACCAGC	TTTACAGCAC	GCTCAAGAGC	ATCCTCCAGC	1860
AGGTGAAGAG	CCATCAAAGC	GCTTGGCCCT	TCATGGAACC	TGTGAAGAGA	ACAGAAGCTC	1920
CAGGATATTA	TGAAGTTATA	AGGTCCCCCA	TGGATCTCAA	AACCATGAGT	GAACGCCTCA	1980
AGAATAGGTA	CTACGTGTCT	AAGAAATTAT	TCATGGCAGA	CTTACAGCGA	GTCTTTACCA	2040
ATTGCAAGA	GTACAACGCC	CCTGAGAGTG	AATACTACAA	ATGTGCCAAT	ATCCTGGAGA	2100
AATTCTTCTT	CAGTAAAATT	AAGGAAGCTG	GATTAATTGA	CAAGTGATTT	TTTTTCCCCC	2160
TCTGCTTCTT	AGAAACTCAC	CAAGCAGTGT	GCCTAAAGCA	AGGT		2204

## (2) INFORMATION FOR SEQ ID NO:11:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2093 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GAATTCCGGC	GAAACCACTC	ATGTCTTTGG	GCGAAGCCTT	CTCCGGTCCA	TTTTCCACCGT	60
TACCCGCCGG	CAGCTGCTGG	AAAAGTTCCG	AGTGGAGAAG	GACAAATTGG	TGCCCCGAGAA	120
GAGGACCCTC	ATCCTCACTC	ACTTCCCCAA	GTAAGGCTCC	TTCTGGCCTA	CCAGGATTTG	180
GCCCCAAGTT	CACATCCTCC	CTGTTGTCCC	CTTTTTTCCA	GGAAGGCTTC	CTGGATTGGT	240
CCCTCCTCTC	CCTCCATGGG	CCTTTTGGGA	TCTGGGCGTC	TACCTGGCAG	ACTTGCCCAT	300
GGCCAGAAG	CAACTTGCTA	GTACTAGTCT	GGGGATGGCA	GATTCCTGTC	CATGCTGGAG	360
GAGGAGATCT	ATGGGGCAAA	CTCTCCAATC	TGGGAGTCAG	GCTTCACCAT	GCCACCCTCA	420

GAGGGGACAC	AGCTGGTTCC	CCGGCCAGCT	TCAGTCAGTG	CAGCGGTTGT	TCCCAGCACC	480
CCCATCTTCA	GCCCCAGCAT	GGGTGGGGGC	AGCAACAGCT	CCCTGAGTCT	GGATTCTGCA	540
GGGGCCGAGC	CTATGCCAGG	CGAGAAGAGG	ACGCTCCCAG	AGAACCTGAC	CCTGGAGGAT	600
GCCAAGCGGC	TCCGTGTGAT	GGGTGACATC	CCCATGGAGC	TGGTCAATGA	GGTCATGCTG	660
ACCATCACTG	ACCCTGCTGC	CATGCTGGGG	CCTGAGACGA	GCCTGCTTTC	GGCCAATGCG	720
GCCCCGGATG	AGACAGCCCG	CCTGGAGGAG	CGCCGCGGCA	TCATCGAGTT	CCATGTCATC	780
GGCAACTCAC	TGACGCCCCA	GGCCAACCGG	CGGGTGTGTC	TGTGGGCTCGT	GGGGCTGCAG	840
AATGTCTTTT	CCCACCAGCT	GCCGCGCATG	CCTAAGGAGT	ATATCGCCCC	CCTCGTCTTT	900
GACCCGAAGC	ACAAGACTCT	GGCCTTGATC	AAGGATGGGC	GGGTCATCGG	TGGCATCTGC	960
TTCCGCATGT	TTCCCACCCA	GGGCTTCACG	GAGATTGTCT	TCTGTGCTGT	CACCTCGAAT	1020
GAGCAGGTCA	AGGGTTATGG	GACCCACCTG	ATGAACCACC	TGAAGGAGTA	TCACATCAAG	1080
CACAACATTC	TCTACTTCCT	CACCTACGCC	GACGAGTACG	CCATCGGCTA	CTTCAAAAAG	1140
CAGGGTTTCT	CCAAGGACAT	CAAGGTGCCC	AAGAGCCGCT	ACCTGGGCTA	CATCAAGGAC	1200
TACGAGGGAG	CGACGCTGAT	GGAGTGTGAG	CTGAATCCCC	GCATCCCCTA	CACGGAGCTG	1260
TCCCACATCA	TCAAGAAGCA	GAAAGAGATC	ATCAAGAAGC	TGATTGAGCG	CAAACAGGCC	1320
CAGATCCGCA	AGGTCTACCC	GGGGCTCAGC	TGCTTCAAGG	AGGGCGTGAG	GCAGATCCCT	1380
GTGGAGAGCG	TTCCTGGCAT	TGCAGAGACA	GGCTGGAAGC	CATTGGGGAA	GGAGAAGGGG	1440
AAGGAGCTGA	AGGACCCCGA	CCAGCTCTAC	ACAACCCTCA	AAAACCTGCT	GGCCCCAATC	1500
AAGTCTCACC	CCAGTGCCTG	GCCCTTCATG	GAGCCTGTGA	AGAAGTCGGA	GGCCCCTGAC	1560
TACTACGAGG	TCATCCGCTT	CCCCATTGAC	CTGAAGACCA	TGACTGAGCG	GCTGCGAAGC	1620
CGCTACTACG	TGACCCGGAA	GCTCTTTGTG	GCCGACCTCG	AGCGGGTCAT	CGCCAACCTG	1680
CGCGAGTACA	ACCCCCCGGA	CAGCGAGTAC	TGCCGCTGTG	CCAGCGCCCT	GGAGAAGTTC	1740
TTCTACTTCA	AGCTCAAGGA	GGGAGGCCCT	ATTGACAAGT	AGGCCCATCT	TTGGGCCGCA	1800
GCCCTGACCT	GGAATGTCTC	CACCTCGGAT	TCTGATCTGA	TCCTTAGGGG	GTGCCCTGGC	1860
CCCACGGACC	CGACTCAGCT	TGAGACACTC	CAGCCAAGGG	TCCTCCGGAC	CCGATCCTGC	1920
AGCTCTTTCT	GGACCTTCAG	GCACCCCCAA	GCCTGCAGCT	CTGTCCCAGC	CTTCACTGTG	1980
TGTGAGAGGT	CTCCTGGGTT	GGGGCCCCAGC	CCCTCTAGAG	TAGCTGGTGG	CCAGGGATGA	2040
ACCTTGCCCA	GCCGTGGTGG	CCCCCAGGCC	TGGTCCCCAA	GAGCCCCGAA	TTC	2093

## (2) INFORMATION FOR SEQ ID NO:12:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9046 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CCTTGTTTGT	GTGCTAGGCT	GGGGGGGAGA	GAGGGCGAGA	GAGAGCGGGC	GAGAGTGGGC	60
AAGCAGGACG	CCGGGCTGAG	TGCTAACTGC	GGGACGCAGA	GAGTGCGGAG	GGGAGTCGGG	120
TCGGAGAGAG	GCGGCAGGGG	CCAGAACAGT	GGCAGGGGGC	CCGGGGCGCA	CCGGGCTGAGG	180
CGACCCCCAG	CCCCCTCCCG	TCCGCACACA	CCCCACCGC	GGTCCAGCAG	CCGGGCCGGC	240
GTCGACGCTA	GGGGGGACCA	TTACATAACC	CGCGCCCCGG	CCGTCTTCTC	CCGCCGCCGC	300
GGCGCCCGAA	CTGAGCCCGG	GGCGGGCGCT	CCAGCACTGG	CCGCCGGCGT	GGGGCGTAGC	360
AGCGCCCGTA	TTATTATTTC	GCGGAAAGGA	AGGCGAAGGA	GGGGAGCGCC	GGCGCGAGGA	420
GGGGCCGCCT	GCGCCCGCCG	CCGGAGCGGG	GCCTCCTCGG	TGGGCTCCGC	GTGCGCGCGG	480
GCGTGCGGGC	GGCGCTGCTC	GGCCCGGCC	CCTCGGCCCT	CTGGTCCGGC	CAGCTCCGCT	540
CCCGGCGTCC	TTGCCGCGCC	TCCGCCGGCC	GCCGCGCGAT	GTGAGGCGGC	GGCGCCAGCC	600
TGGCTCTCGG	CTCGGGCGAG	TTCTCTGCGG	CCATTAGGGG	CCGGTGCGGC	GGCGGCGCGG	660
AGCGCGGCGG	CAGGAGGAGG	GTTCGGAGGG	TGGGGGCGCA	GGCCCCGGAG	GGGCCACCGG	720
GAGGAGGTGA	GTGTCTCTTG	TGCGCTCCTC	CTCTCCCCC	TTTTCGCCCC	CGCCTCCTTG	780
TGGCGATGAG	AAGGAGGAGG	ACAGCGCCGA	GGAGGAAGAG	GTTGATGGCG	GCGGCGGAGC	840
TCCGAGAGAC	CTCGGCTGGG	CAGGGGCCGG	CCGTGGCGGG	CCGGGGACTG	CGCCTCTAGA	900
GCCGCGAGTT	CTCGGGAATT	CGCCGCAGCG	GACCGGCCTC	GGCGAATTTG	TGCTCTTGTG	960
CCCTCCTCCG	GGCTTGGGCC	AGGCCGGCCC	CTCGCACTTG	CCCTTACCTT	TTCTATCGAG	1020
TCCGCATCCC	TCTCCAGCCA	CTGCGACCCG	GCGAAGAGAA	AAAGGAACTT	CCCCACCCC	1080
CTCGGGTGCC	CTCGGAGCCC	CCCAGCCAC	CCCTGGGTGC	GGCGCGGGGA	CCCCGGGCCG	1140
AAGAAGAGAT	TTCCTGAGGA	TTCTGGTTTT	CCTCGCTTGT	ATCTCCGAAA	GAATTAAAAA	1200
TGGCCGAGAA	TGTGGTGGAA	CCGGGGCCGC	CTTCAGCCAA	GCGGCCTAAA	CTCTCATCTC	1260
CGGCCCTCTC	GGCGTCCGCC	AGCGATGGCA	CAGATTTTGG	CTCTCTATTT	GACTTGGAGC	1320
ACGACTTACC	AGATGAATTA	ATCAACTCTA	CAGAATTGGG	ACTAACCAAT	GGTGGTGATA	1380

TTAATCAGCT	TCAGACAAGT	CTTGGCATGG	TACAAGATGC	AGCTTCTAAA	CATAAACAGC	1440
TGTCAGAATT	GCTGCGATCT	GGTAGTTCCC	CTAACCTCAA	TATGGGAGTT	GGTGGCCAG	1500
GTCAAGTCAT	GGCCAGCCAG	GCCCAACAGA	GCAGTCCTGG	ATTAGGTTTG	ATAAATAGCA	1560
TGGTCAAAAG	CCCAATGACA	CAGGCAGGCT	TGACTTCTCC	CAACATGGGG	ATGGGCACTA	1620
GTGGACCAAA	TCAGGGTCCT	ACGCAGTCAA	CAGGTATGAT	GAACAGTCCA	GTAAATCAGC	1680
CTGCCATGGG	AATGAACACA	GGGACGAATG	CGGGCATGAA	TCCTGGAATG	TTGGCTGCAG	1740
GCAATGGACA	AGGGATAATG	CCTAATCAAG	TCATGAACGG	TTCAATTGGA	GCAGGCCGAG	1800
GGCGACAGGA	TATGCAGTAC	CCAAACCCAG	GCATGGGAAG	TGCTGGCAAC	TTACTGACTG	1860
AGCCTCTTCA	GCAGGGCTCT	CCCCAGATGG	GAGGACAAAC	AGGATTGAGA	GGCCCCCAGC	1920
CTCTTAAGAT	GGGAATGATG	AACAACCCCA	ATCCTTATGG	TTCAACCATAT	ACTCAGAATC	1980
CTGGACAGCA	GATTGGAGCC	AGTGGCCTTG	GTCTCCAGAT	TCAGACAAAA	ACTGTACTAT	2040
CAAATAACTT	ATCTCCATTT	GCTATGGACA	AAAAGGCAGT	TCCTGGTGGA	GGAATGCCCA	2100
ACATGGGTCA	ACAGCCAGCC	CCGCAGGTCC	AGCAGCCAGG	TCTGGTGACT	CCAGTTGCCC	2160
AAGGGATGGG	TTCTGGAGCA	CATACAGCTG	ATCCAGAGAA	GCGCAAGCTC	ATCCAGCAGC	2220
AGCTTGTCTT	CCTTTTGCAT	GCTCACAACT	GCCAGCGCCG	GGAACAGGCC	AATGGGGAAG	2280
TGAGGCAGTG	CAACCTTCCC	CACTGTGCGA	CAATGAAGAA	TGTCCTAAAC	CACATGACAC	2340
ACTGCCAGTC	AGGCAAGTCT	TGCCAAGTGG	CACACTGTGC	ATCTTCTCGA	CAAATCATTT	2400
CACACTGGAA	GAATTGTACA	AGACATGATT	GTCTGTGTG	TCTCCCCCTC	AAAAATGCTG	2460
GTGATAAGAG	AAATCAACAG	CCAATTTTGA	CTGGAGCACC	CGTTGGACTT	GGAAATCCTA	2520
GCTCTCTAGG	GGTGGGTCAA	CAGTCTGCCC	CCAACCTAAG	CACTGTTAGT	CAGATTGATC	2580
CCAGCTCCAT	AGAAAGAGCC	TATGCAGCTC	TTGGACTACC	CTATCAAGTA	AATCAGATGC	2640
CGACACAACC	CCAGGTGCAA	GCAAAGAACC	AGCAGAATCA	GCAGCCTGGG	CAGTCTCCCC	2700
AAGGCATGCG	GCCCATGAGC	AACATGAGTG	CTAGTCCTAT	GGGAGTAAAT	GGAGGTGTAG	2760
GAGTTCAAAC	GCCGAGTCTT	CTTTCTGACT	CAATGTTGCA	TTCAAGCCATA	AATTCTCAAA	2820
ACCCAATGAT	GAGTGAAAAT	GCCAGTGTGC	CCTCCCTGGG	TCCTATGCCA	ACAGCAGCTC	2880
AACCATCCAC	TACTGGAATT	CGGAAACAGT	GGCAGGAAGA	TATTACTCAG	GATCTTCGAA	2940
ATCATCTTGT	TACAAAATC	GTCCAAGCCA	TATTTCTTAC	GCCGGATCCT	GCTGCTTTAA	3000
AAGACAGACG	GATGGAAAAC	CTAGTTGCAT	ATGCTCGGAA	AGTTGAAGGG	GACATGTATG	3060
AATCTGCAAA	CAATCGAGCG	GAATACTACC	ACCTTCTAGC	TGAGAAAATC	TATAAGATCC	3120
AGAAAGAACT	AGAAGAAAAA	CGAAGGACCA	GACTACAGAA	GCAGAACATG	CTACCAAATG	3180
CTGCAGGCAT	GGTTCCAGTT	TCCATGAATC	CAGGGCCTAA	CATGGGACAG	CCGCAACCAG	3240
GAATGACTTC	TAATGGCCCT	CTACCTGACC	CAAGTATGAT	CCGTGGCAGT	GTGCCAAACC	3300
AGATGATGCC	TGCAATAACT	CCACAATCTG	GTTTGAATCA	ATTTGGCCAG	ATGAGCATGG	3360
CCCAGCCCCC	TATTGTACCC	CGGCAAAACC	CTCCTCTTCA	GCACCATGGA	CAGTTGGCTC	3420
AACCTGGAGC	TCTCAACCCG	CCTATGGGCT	ATGGGCCTCG	TATGCAACAG	CCTTCCAACC	3480
AGGGCCAGTT	CCTTCCTCAG	ACTCAGTTCC	CATCACAGGG	AATGAATGTA	ACAAATATCC	3540
CTTTGGCTCC	GTCCAGCGGT	CAAGCTCCAG	TGTCCTAAGC	ACAAATGTCT	AGTTCTTCTT	3600
GCCCCGTGAA	CTCTCCTATA	ATGCCTCCAG	GGTCTCAGGG	GAGCCACATT	CAGTGTCCCC	3660
AGCTTCCTCA	ACCAGTCTT	CATCAGAATT	CACCTCGCC	TGTACCTAGT	CGTACCCCCA	3720
CCCCTCACCA	TACTCCCCCA	AGCATAGGGG	CTCAGCAGCC	ACCAGCAACA	ACAATTCCAG	3780
CCCCTGTTCC	TACACCACCA	GCCATGCCAC	CTGGGCCACA	GTCCCAGGCT	CTACATCCCC	3840
CTCCAAGGCA	GACACCTACA	CCACCAACAA	CACAACCTCC	CCAACAAGTG	CAGCCTTCAC	3900
TTCTGCTGTC	ACCTTCTGCT	GACCAGCCCC	AGCAGCAGCC	TCGCTCACAG	CAGAGCACAG	3960
CAGCGTCTGT	TCCTACCCCA	AACGCACCGG	TGCTTCCCTC	GCAGCCTGCA	ACTCCACTTT	4020
CCCAGCCAGC	TGTAAGCATT	GAAGGACAGG	TATCAAATCC	TCCATCTACT	AGTAGCACAG	4080
AAGTGAATTC	TCAGGCCATT	GCTGAGAAGC	AGCCTTCCCA	GGAAGTGAAG	ATGGAGGCCA	4140
AAATGGAAGT	GGATCAACCA	GAACCAGCAG	ATACGCAGCC	GGAGGATATT	TCAGAGTCTA	4200
AAGTGGAAGA	CTGTAAATATG	GAATCTACCG	AAACAGAAGA	GAGAAGCACT	GAGTTAAAAA	4260
CTGAAATAAA	AGAGGAGGAA	GACCAGCCAA	GTACTTCAGC	TACCCAGTCA	TCTCCGGCTC	4320
CAGGACAGTC	AAAGAAAAAG	ATTTTCAAAC	CAGAAGAACT	ACGACAGGCA	CTGATGCCAA	4380
CATTGGAGGC	ACTTTACCGT	CAGGATCCAG	AATCCCTTCC	CTTTTCGTCAA	CCTGTGGACC	4440
CTCAGCTTTT	AGGAATCCCT	GATTACTTTG	ATATTGTGAA	GAGCCCCATG	GATCTTTCTA	4500
CCATTAAGAG	GAAGTTAGAC	ACTGGACAGT	ATCAGGAGCC	CTGGCAGTAT	GTGATGATA	4560
TTTGGCTTAT	GTTCAATAAT	GCCTGGTTAT	ATAACCGGAA	AACATCACGG	GTATACAAAT	4620
ACTGCTCCAA	GCTCTCTGAG	GTCTTTGAAC	AAGAAATTGA	CCCAGTGATG	CAAAGCCTTG	4680
GATACTGTTG	TGGCAGAAAAG	TTGGAGTTCT	CTCCACAGAG	ACTGTGTTGC	TACGGCAAAC	4740
AGTTGTGCAC	AATACCTCGT	GATGCCACTT	ATTACAGTTA	CCAGAACAGG	TATCATTTCT	4800
GTGAGAAGTG	TTTCAATGAG	ATCCAAGGGG	AGAGCGTTTC	TTTGGGGGAT	GACCCCTCCC	4860
AGCCTCAAAC	TACAATAAAT	AAAGAACAAT	TTTCCAAGAG	AAAAAATGAC	ACACTGGATC	4920
CTGAAGTGT	TGTTGAATGT	ACAGAGTGCG	GAAGAAAGAT	GCATCAGATC	TGTGTCTTTC	4980
ACCATGAGAT	CATCTGGCCT	GCTGGATTCT	TCTGTGATGG	CTGTTTAAAG	AAAAGTGCAC	5040
GAAGTAGGAA	AGAAAAATAAG	TTTTCTGCTA	AAAGGTTGCC	ATCTACCAGA	CTTGGCACCT	5100
TTCTAGAGAA	TCGTGTGAAT	GACTTTCCTGA	GGCGACAGAA	TCACCCTGAG	TCAGGAGAGG	5160

TCAGTGTAG	AGTAGTTCAT	GCTTCTGACA	AAACCGTGGA	AGTAAAACCA	GGCATGAAAG	5220
CAAGGTTTGT	GGACAGTGGA	GAGATGGCAG	AATCCTTTCC	ATACCGAACC	AAAGCCCTCT	5280
TTGCCCTTGA	AGAAATTGAT	GGTGTGACC	TGTGCTTCTT	TGGCATGCAT	GTTCAGAGT	5340
ATGGCTCTGA	CTGCCCTCCA	CCCAACCAGA	GGAGAGTATA	CATATCTTAC	CTCGATAGTG	5400
TTCATTTCTT	CCGTCCTAAA	TGCTTGAGGA	CTGCAGTCTA	TCATGAAATC	CTAATTGGAT	5460
ATTTAGAATA	TGTCAAGAAA	TTAGGTTACA	CAACAGGGCA	TATTTGGGCA	TGTCCACCAA	5520
GTGAGGGAGA	TGATTATATC	TTCCATTGCC	ATCCTCCTGA	CCAGAAGATA	CCCAAGCCCA	5580
AGCGACTGCA	GGAATGGTAC	AAAAAATGC	TTGACAAGGC	TGTATCAGAG	CGTATTGTCC	5640
ATGACTACAA	GGATATTTTT	AAACAAGCTA	CTGAAGATAG	ATTAACAAGT	GCAAAGGAAT	5700
TGCCTTATTT	CGAGGGTGAT	TTCTGGCCCA	ATGTTCTGGA	AGAAAAGCATT	AAGGAACTGG	5760
AACAGGAGGA	AGAAGAGAGA	AAACGAGAGG	AAAACACCAG	CAATGAAAGC	ACAGATGTGA	5820
CCAAGGGAGA	CAGCAAAAT	GCTAAAAAGA	AGAATAATAA	GAAAACCAGC	AAAAATAAGA	5880
GCAGCCTGAG	TAGGGGCAAC	AAGAAGAAAC	CCGGGATGCC	CAATGTATCT	AACGACCTCT	5940
CACAGAAACT	ATATGCCACC	ATGGAGAAGC	ATAAAGAGGT	CTTCTTTGTG	ATCCGCCTCA	6000
TTGCTGGCCC	TGCTGCCAAC	TCCCTGCCTC	CCATTGTTGA	TCCTGATCCT	CTCATCCCCT	6060
CGCATCTGAT	GGATGGTCGG	GATGCGTTTC	TCACGTGGC	AAGGGACAAG	CACCTGGAGT	6120
TCTCTTCACT	CCGAAGAGCC	CAGTGGTCCA	CCATGTGCAT	GCTGGTGGAG	CTGCACACGC	6180
AGAGCCAGGA	CCGCTTTGTC	TACACCTGCA	ATGAATGCAA	GCACCATGTG	GAGACACGCT	6240
GGCACTGTAC	TGTCTGTGAG	GATTATGACT	TGTGTATCAC	CTGCTATAAC	ACTAAAAACC	6300
ATGACCACAA	AATGGAGAAA	CTAGGCCTTG	GCTTAGATGA	TGAGAGCAAC	AACCAGCAGG	6360
CTGCAGCCAC	CCAGAGCCCA	GGCGATTCTC	GCCGCCTGAG	TATCCAGCGC	TGCATCCAGT	6420
CTCTGGTCCA	TGCTTGCCAG	TGTCGGAATG	CCAATTGCTC	ACTGCCATCC	TGCCAGAAGA	6480
TGAAGCGGST	TGTGCAGCAT	ACCAAGGGTT	GCAAACGGAA	AACCAATGGC	GGGTGCCCCA	6540
TCTGCAAGCA	GCTCATTGCC	CTCTGCTGCT	ACCATGCCAA	GCACTGCCAG	GAGAACAAAT	6600
GCCCGGTGCC	GTTCTGCCTA	AACATCAAGC	AGAAGCTCCG	GCAGCAACAG	CTGCAGCACC	6660
GACTACAGCA	GGCCCAAATG	CTTCGCAGGA	GGATGGCCAG	CATGCAGCGG	ACTGGTGTGG	6720
TTGGGCGAGCA	ACAGGGCCTC	CCTTCCCCCA	CTCCTGCCAC	TCCAACGACA	CCAATGGCC	6780
AACAGCCAAC	CACCCGCGAG	ACGCCCCAGC	CCACTTCTCA	GCCTCAGCCT	ACCCCTCCCA	6840
ATAGCATGCC	ACCCTACTTG	CCCAGGACTC	AAGTGTCTGG	CCCTGTGTCC	CAGGGTAAGG	6900
CAGCAGGCCA	GGTGACCCCT	CCAACCCCTC	CTCAGACTGC	TCAGCCACCC	CTTCCAGGGC	6960
CCCCACCTAC	AGCAGTGGA	ATGGCAATGC	AGATTCAGAG	AGCAGCGGAG	ACGCAGCGCC	7020
AGATGGCCCA	CGTGCAAATT	TTTCAAAGGC	CAATCCAACA	CCAGATGCCC	CCGATGACTC	7080
CCATGGCCCC	CATGGGTATG	AACCCACCTC	CCATGACCAG	AGGTCCCAGT	GGGCATTTGG	7140
AGCCAGGGAT	GGGACCGACA	GGGATGCAGC	AACAGCCACC	CTGGAGCCAA	GGAGGATTGC	7200
CTCAGCCCCA	GCAACTACAG	TCTGGGATGC	CAAGGCCAGC	CATGATGTCA	GTGGCCAGC	7260
ATGGTCAACC	TTTGAACATG	GCTCCACAAC	CAGGATTGGG	CCAGGTAGGT	ATCAGCCCAC	7320
TCAAACCAGG	CACTGTGTCT	CAACAAGCCT	TACAAAACCT	TTTGCGGACT	CTCAGGTCTC	7380
CCAGCTCTCC	CCTGCAGCAG	CAACAGGTGC	TTAGTATCCT	TCACGCCAAC	CCCCAGCTGT	7440
TGGCTGCATT	CATCAAGCAG	CGGGCTGCCA	AGTATGCCAA	CTCTAATCCA	CAACCCATCC	7500
CTGGGCGAGC	TGGCATGCCC	CAGGGGCAGC	CAGGGCTACA	GCCACCTACC	ATGCCAGGTC	7560
AGCAGGGGGT	CCACTCCAAT	CCAGCCATGC	AGAATCTGAA	TCCAATGCAG	GCGGGCGTTC	7620
AGAGGGCTGG	CCTGCCCCAG	CAGCAACCAC	AGCAGCAACT	CCAGCCACCC	ATGGGAGGGA	7680
TGAGCCCCCA	GGCTCAGCAG	ATGAACATGA	ACCACAACAC	CATGCCTTCA	CAATTCCGAG	7740
ACATCTTGAG	ACGACAGCAA	ATGATGCAAC	AGCAGCAGCA	ACAGGGAGCA	GGGCCAGGAA	7800
TAGGCCCTGG	AATGGCCAAC	CATAACCAGT	TCCAGCAACC	CCAAGGAGTT	GGCTACCCAC	7860
CACAGCCGCA	GCAGCGGATG	CAGCATCACA	TGCAACAGAT	GCAACAAGGA	AATATGGGAC	7920
AGATAGGCCA	GCTTCCCCAG	GCCTTGGGAG	CAGAGGCAGG	TGCCAGTCTA	CAGGCCTATC	7980
AGCAGCGACT	CCTTCAGCAA	CAGATGGGGT	CCCCTGTTCA	GCCCAACCCC	ATGAGCCCCC	8040
AGCAGCATAT	GCTCCCAAAT	CAGGCCAGT	CCCCACACCT	ACAAGGCCAG	CAGATCCCTA	8100
ATTCTCTCTC	CAATCAAGTG	CGCTCTCCCC	AGCCTGTCCC	TTCTCCACGG	CCACAGTCCC	8160
AGCCCCCCCC	CTCCAGTCCT	TCCCCAAGGA	TGCAGCCTCA	GCCTTCTCCA	CACCACGTTT	8220
CCCCACAGAC	AAGTTCCCCA	CATCCTGGAC	TGGTAGCTGC	CCAGGCCAAC	CCATGGAAC	8280
AAGGGCATTT	TGCCAGCCCG	GACCAGAATT	CAATGCTTTC	TCAGCTTGCT	AGCAATCCAG	8340
GCATGGCAAA	CCTCCATGGT	GCAAGCGCCA	CGGACCTGGG	ACTCAGCACC	GATAACTCAG	8400
ACTTGAATTC	AAACCTCTCA	CAGAGTACAC	TAGACATACA	CTAGAGACAC	CTTGATTTTT	8460
GGGAGCAAAA	AAATTATTTT	CTCTTAACAA	GACTTTTTGT	ACTGAAAACA	ATTTTTTTGA	8520
ATCTTTCGTA	GCCTAAAAGA	CAATTTTCCT	TGGAACACAT	AAGAAGTGTG	CAGTAGCCGT	8580
TTGTGGTTTA	AAGCAAACAT	GCAAGATGAA	CCTGAGGGAT	GATAGAATAC	AAAGAATATA	8640
TTTTTGTTAT	GGGCTGGTTA	CCACCAGCCT	TTCTTCCCCT	TTGTGTGTGT	GGTTCAAGTG	8700
TGCACTGGGA	GGAGGCTGAG	GCCTGTGAAG	CCAAACAATA	TGCTCCTGCC	TTGCACCTCC	8760
AATAGGTTTT	ATTATTTTTT	TTAAATTAAT	GAACATATGT	AATATTAATG	AACATATGTA	8820
ATATTAATAG	TTATTATTTA	CTGGTGACAG	TGGTTGACAT	TTTTCCCTAT	TTTCCCTACT	8880
TTATGGAAGA	GTAAAAACAT	TTCTAAACCA	GAGGACAAAA	GGGGTTAATG	TTACTTTGAA	8940

ATTACATTCT ATATATATAT AAATATATAT AAATATATAT TAAAATACCA GTTTTTTTTC 9000  
TCTGGGTGCA AAGATGTTCA TTCTTTTAAA AAATGTTTAA AAAAAA 9046

## (2) INFORMATION FOR SEQ ID NO:13:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7326 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

ATGGCCGAGA ACTTGCTGGA CGGACCGCCC AACCCCAAAC GAGCCAAACT CAGCTCGCCC 60  
GGCTTCTCCG CGAATGACAA CACAGATTTT GGATCATTGT TTGACTTGGA AAATGACCTT 120  
CCTGATGAGC TGATCCCCAA TGGAGAATTA AGCCTTTTAA ACAGTGGGAA CCTTGTTCCA 180  
GATGCTGCGT CCAAACATAA ACAACTGTCA GAGCTTCTTA GAGGAGGCAG CGGCTCTAGC 240  
ATCAACCCAG GGATAGGCAA TGTGAGTGCC AGCAGCCCTG TGCAACAGGG CCTTGGTGCC 300  
CAGGCTCAGG GGCAGCCGAA CAGTACAAAC ATGGCCAGCT TAGGTGCCAT GGGCAAGAGC 360  
CCTCTGAACC AAGGAGACTC ATCAACACCC AACCTGCCCC AACAGGCAGC CAGCACCTCT 420  
GGGCCCCACT CCCCTGCCTC CCAAGCACTG AATCCACAAG CACAAAAGCA AGTAGGGCTG 480  
GTGACCAGTA GTCCTGCCAC ATCACAGACT GGACCTGGGA TCTGCATGAA TGCTAACTTC 540  
AACCAGACCC ACCCAGGCCT TCTCAATAGT AACTCTGGCC ATAGCTTAAT GAATCAGGCT 600  
CAACAAGGGG AAGCTCAAGT CATGAATGGA TCTCTTGGGG CTGCTGGAAG AGGAAGGGGA 660  
GCTGGAATGC CCTACCCTGC TCCAGCCATG CAGGGGGCCA CAAGCAGTGT GGTGCGGAG 720  
ACCTTGACAC AGGTTTCCCC ACAAAATGGCT GGCCATGCTG GACTAAATAC AGCACAGGCA 780  
GGAGGCATGA CCAAGATGGG AATGACTGGT ACCACAAGTC CATTTGGACA ACCCTTTAGT 840  
CAAACCTGGG GGCAGCAGAT GGGAGCCACT GGAGTGAACC CCCAGTTAGC CAGCAAACAG 900  
AGCATGGTCA ATAGTTTACC TGCTTTTCCT ACAGATATCA AGAATACTTC AGTCACCACT 960  
GTGCCAAATA TGTCCCAGTT GCAAACATCA GTGGGAATTG TACCCACACA AGCAATTGCA 1020  
ACAGGCCCCA CAGCAGACCC TGAAAAACGC AAACCTGATAC AGCAGCAGCT GGTTCTACTG 1080  
CTTCATGCCC ACAAAATGTCA GAGACGAGAG CAAGCAAATG GAGAGGTTTC NGCCTGTTCT 1140  
CTCCCACT GTCGAACCAT GAAAAACGTT TTGAATCACA TGACACATTG TCAGGCTCCC 1200  
AAAGCCTGCC AAGTTGCCCA TTGTGCATCT TCACGACAAA TCATCTCTCA TTGGAAGAAC 1260  
TGCACACGAC ATGACTGTCC TGTTTGCCCT CTTTGAAAA ATGCCAGTGA CAAGCGAAAC 1320  
CAACAAACCA TCCTGGGATC TCCAGCTAGT GGAATTCAAA ACACAATTGG TTCTGTTGGT 1380  
GCAGGGCAAC AGAATGCCAC TTCCTTAAGT AACCCAAATC CCATAGACCC CAGTCCCATG 1440  
CAGCGGGCCT ATGCTGCTCT AGGACTCCCC TACATGAACC AGCCTCAGAC GCAGCTGCAG 1500  
CCTCAGGTTT CTGGCCAGCA ACCAGCACAG CCTCCAGCCC ACCAGCAGAT GAGGACTCTC 1560  
AATGCCCTAG GAAACAACCC CATGAGTGTC CCAGCAGGAG GAATAACAAC AGATCAACAG 1620  
CCACCAAAC TGATTTTACA ATCAGCTCTT CCAACTTCCT TGGGGGCTAC CAATCCACTG 1680  
ATGAATGATG GTTCAAACCT TGGTAACATT GGAAGCCTCA GCACGATACC TACAGCAGCG 1740  
CTCCTTCCA GCACTGGTGT TCGAAAAGGC TGGCATGAAC ATGTGACTCA GGACCTCAGG 1800  
AGTCATCTAG TCCATAAACT CGTTCAGGCC ATCTTCCCAA CTCCAGACCC TGCAGCTCTG 1860  
AAAGATCGCC GCATGGAGAA CCTGGTTGCC TATGCTAAGA AAGTGGAGGG AGACATGTAT 1920  
GAGTCTGCTA ATAGCAGGGA TGAATACTAT CATTTATTAG CAGAGAAAAT CTATAAAATA 1980  
CAAAAAGAAC TAGAAGAAAA GCGGAGGACA CGTTTACATA AGCAAGGCAT CCTGGGTAAC 2040  
CAGCCAGCTT TACCAGCTTC TGGGGCTCAG CCCCTGTGA TTCCACCAGC CCACTCTGTA 2100  
AGACCTCAA ATGGGGCCCT GCCTTTGCCA GTGAATCGCA TGCAGTTTC TCAAGGGATG 2160  
AATTCAATTA ACCCAATGTC CCTGGGAAAC GTCCAGTTGC CACAGGCACC CATGGGACCT 2220  
CGTGCAGCCT CCCCTATGAA CCACTCTGTG CAGATGAACA GCATGGCCTC AGTTCCGGGT 2280  
ATGGCCATTT CTCCTTCACG GATGCCTCAG CCTCCAAATA TGATGGGCAC TCATGCCAAC 2340  
AACATTATGG CCCAGGCACC TACTCAGAAC CAGTTTCTGC CACAGAACCA GTTTCCATCA 2400  
TCCAGTGGGG CAATGAGTGT GAACAGTGTG GGCATGGGGC AAACAGCAGC CCAGGCAGGT 2460  
GTTTCACAGG GTCAGGAACC TGGAGCTGCT CTCCTTAACC CTCTGAACAT GCTGGCACCC 2520  
CAGGCCAGCC AGCTGCCTTG CCCACCAGTG ACACAGTCAC CATTCACACC GACTCCACCT 2580  
CCTGCTTCCA CAGCTGCTGG CATGCCCTCT CTCCAACATC CAACGGCACC AGGAATGACC 2640  
CCTCCTCAGC CAGCAGCTCC CACTCAGCCA TCTACTCCTG TGTCATCTGG GCAGACTCCT 2700  
ACCCCAACTC CTGGCTCAGT GCCCAGCGCT GCCCAAACAC AGAGTACCCC TACAGTCCAG 2760  
GCAGCAGCAC AGGCTCAGGT GACTCCACAG CCTCAGACCC CAGTGCAGCC ACCATCTGTG 2820  
GCTACTCCTC AGTCATCACA GCAGCAACCA ACGCCTGTGC ATACTCAGCC ACCTGGCACA 2880  
CCGCTTTCTC AGGCAGCAGC CAGCATTGAT AATAGAGTCC CTACTCCCTC CACTGTGACC 2940

AGTGCTGAAA	CCAGTTCCCA	GCAGCCAGGA	CCCGATGTGC	CCATGCTGGA	AATGAAGACA	3000
GAGGTGCAGA	CAGATGATGC	TGAGCCTGAA	CCTACTGAAT	CCAAGGGGGA	ACCTCGGTCT	3060
GAGATGATGG	AAGAGGATTT	ACAAGGTTCT	TCCCAAGTAA	AAGAAGAGAC	AGATACGACA	3120
GAGCAGAAGT	CAGAGCCAAT	GGAAGTAGAA	GAAAAGAAAC	CTGAAGTAAA	AGTGGAAGCT	3180
AAAGAGGAAG	AAGAGAACAG	TTCGAACGAC	ACAGCCTCAC	AATCAACATC	TCCTTCCCAG	3240
CCACGCAAAA	AAATCTTTAA	ACCCGAGGAG	CTACGCCAGG	CACTTATGCC	AACTCTAGAA	3300
GCACTCTATC	GACAGGACCC	AGAGTCTTTG	CCTTTTCGTC	AGCCTGTAGA	TCCTCAGCTC	3360
CTAGGAATCC	CAGATTATTT	TGATATAGTG	AAGAATCCTA	TGGACCTTTC	TACCATCAAA	3420
CGAAAGCTGG	ACACAGGGCA	ATATCAAGAA	CCCTGGCAGT	ATGTGGATGA	TGTCAGGCTT	3480
ATGTTCAACA	ATGCGTGGCT	ATATAATCGT	AAAACGTCCC	GTGTATATAA	ATTTTGCAGT	3540
AAACTTGCAG	AGGTCTTTGA	ACAAGAAATT	GACCCTGTCA	TGCAGTCTCT	TGGATATTGC	3600
TGTGGACGAA	AGTATGAGTT	CTCCCCACAG	ACTTTGTGCT	GTTACGGAAA	GCAGCTGTGT	3660
ACAATTCCTC	GTGATGCAGC	CTACTACAGC	TATCAGAATA	GGTATCATTT	CTGTGGGAAG	3720
TGTTTTCACAG	AGATCCAGGG	CGAGAATGTG	ACCCTGGGTG	ACGACCCTTC	CCAACCTCAG	3780
ACGACAATTT	CCAAGGATCA	ATTTGAAAAG	AAGAAAAATG	ATACCTTAGA	TCCTGAACCT	3840
TTTGTGTGACT	GCAAAGAGTG	TGGCCGGAAG	ATGCATCAGA	TTTGTGTTCT	ACACTATGAC	3900
ATCATTTGGC	CTTCAGGTTT	TGTGTGTGAC	AACTGTTTGA	AGAAAACTGG	CAGACCTCGG	3960
AAAGAAAACA	AATTCAGTGC	TAAGAGGCTG	CAGACCACAC	GATTGGGAAA	CCACTTAGAA	4020
GACAGAGTGA	ATAAGTTTTT	GCGGCGCCAG	AATCACCCTG	AAGCTGGGGA	GGTTTTTGTG	4080
AGAGTGGTGG	CCAGCTCAGA	CAAGACTGTG	GAGGTCAAGC	CGGGAATGAA	GTCAAGGTTT	4140
GTGGATTCTG	GAGAGATGTC	GGAATCTTTC	CCATATCGTA	CCAAAGCACT	CTTTGCTTTT	4200
GAGGAGTCTG	ATGGAGTCGA	TGTGTGCTTT	TTTGGGATGC	ATGTGCAAGA	TACGGCTCTG	4260
ATTGCCCCCC	ACCAAATACA	AGGCTGTGTA	TACATATCTT	ATCTGGACAG	TATTCAATTC	4320
TTCCGGCCCC	GCTGCCTCCG	GACAGCTGTT	TACCATGAGA	TCCTCATCGG	ATATCTCGAG	4380
TATGTGAAGA	AATTGGTGTA	TGTGACAGCA	CATATTTGGG	CCTGTCCCCC	AAGTGAAGGA	4440
GATGACTATA	TCTTTCATTG	CCACCCCCCT	GACCAGAAAA	TCCCCAAACC	AAAACGACTA	4500
CAGGAGTGGT	ACAAGAAGAT	GCTGGACAAG	GCGTTTGCA	AGAGGATCAT	TAACGACTAT	4560
AAGGACATCT	TCAAACAAGC	GAACGAAGAC	AGGCTCACGA	GTGCCAAGGA	GTTGCCCTAT	4620
TTTGAAGGAG	ATTTCTGGCC	TAATGTGTTG	GAAGAAAGCA	TTAAGGAACT	AGAACAAGAA	4680
GAAGAAGAAA	GGAAAAAAGA	AGAGAGTACT	GCAGCGAGTG	AGACTCCTGA	GGGCAGTCAG	4740
GGTGACAGCA	AAAATGCGAA	GAAAAAGAAC	AACAAGAAGA	CCAACAAAAA	CAAAAGCAGC	4800
ATTAGCCGCG	CCAACAAGAA	GAAGCCCAGC	ATGCCCAATG	TTTCCAACGA	CCTGTCGCAG	4860
AAGCTGTATG	CCACCATGGA	GAAGCACAAG	GAGGTATTCT	TTGTGATTCA	TCTGCATGCT	4920
GGCCTGTGTA	TCAGCACTCA	GCCCCCATC	TGGGACCCTG	ATCCTCTGCT	TAGCTGTGAC	4980
CTCATGGATG	GGCGAGATGC	CTTCCTCACC	CTGGCCAGAG	ACAAGCACTG	GGAATTCTCT	5040
TCCTTACGCC	GCTCCAAATG	GTCCACTCTG	TGCATGCTGG	TGGAGCTGCA	CACACAGGGC	5100
CAGGACCGCT	TTGTTTATAC	CTGCAATGAG	TGCAACACCC	ATGTGGAAAC	ACGCTGGCAC	5160
TGCACTGTGT	GTGAGGACTA	TGACCTTTGT	ATCAATTGCT	ACAACACAAA	GAGCCACACC	5220
CATAAGATGG	TGAAGTGGGG	GCTAGGCCTA	GATGATGAGG	GCAGCAGTCA	GGGTGAGCCA	5280
CAGTCCAAGA	GCCCCCAGGA	ATCCCGGCGT	CTCAGCATCC	AGCGCTGCAT	CCAGTCCCTG	5340
GTGCATGCCT	GCCAGTGTGC	CAATGCCAAC	TGCTCACTGC	CGTCTTGCCA	GAAGATGAAG	5400
CGAGTCGTGC	AGCACACCAA	GGGCTGCAAG	CGCAAGACTA	ATGGAGGATG	CCCAGTGTGC	5460
AAGCAGCTCA	TTGCTCTTTG	CTGCTACCAC	GCCAAACACT	GCCAAGAAAA	TAAATGCCCT	5520
GTGCCCTTCT	GCCTCAACAT	CAAACATAAC	GTCCGCCAGC	AGCAGATCCA	GCACTGCCTG	5580
CAGCAGGCTC	AGCTCATGCG	CCGGCGAATG	GCAACCATGA	ACACCCGCAA	TGTGCCTCAG	5640
CAGAGTTTGA	CTTCTCCTAC	CTCAGCACC	CCCGGGACTC	CTACACAGCA	GCCCAGCACA	5700
CCCCAAACAC	CACAGCCCCC	AGCCCAGCCT	CAGCCTTCAC	CTGTTAACAT	GTCACCAGCA	5760
GGCTTCCCTA	ATGTAGCCCG	GACTCAGCCC	CCAACAATAG	TGTCTGCTGG	GAAGCCTACC	5820
AACCAGGTGC	CAGCTCCCCC	ACCCCCCTGC	CAGCCCCCAC	CTGCAGCAGT	AGAAGCAGCC	5880
CGGCAAAATG	AACGTGAGGC	CCAGCAGCAG	CAGCACCTAT	ACCGAGCAAA	CATCAACAAT	5940
GGCATGCCCC	CAGGACGTGA	CGGTATGGGG	ACCCAGGAA	GCCAAATGAC	TCCTGTGGGC	6000
CTGAATGTGC	CCCGTCCCAA	CCAAGTCAGT	GGGCCTGTCA	TGTCTAGTAT	GCCACCTGGG	6060
CAGTGGCAGC	AGGCACCCAT	CCCTCAGCAG	CAGCCGATGC	CAGGCATGCC	CAGGCCTGTA	6120
ATGTCCATGC	AGGCCCAGGC	AGCAGTGGCT	GGGCCACGGA	TGCCCCAATGT	GCAGCCAAAC	6180
AGGAGCATCT	CGCCAAGTGC	CCTGCAAGAC	CTGCTACGGA	CCCTAAAGTC	ACCCAGCTCT	6240
CCTCAGCAGC	AGCAGCAGGT	GCTGAACATC	CTTAAATCAA	ACCCACAGCT	AATGGCAGCT	6300
TTCATCAAAC	AGCGCACAGC	CAAGTATGTG	GCCAATCAGC	CTGGCATGCA	GCCCCAGCCC	6360
GGACTTCAAT	ACCCGCTGG	TATGCAGCCC	CAGCCTGGCA	TGCACCAGCA	GCCTAGTTTG	6420
CAAAACCTGA	ACGCAATGCA	AGCTGGTGTG	CCACGGCCTG	GTGTGCCTCC	ACCACAACCA	6480
GCAATGGGAG	GCCTGAATCC	CCAGGGACAA	GCTCTGAACA	TCATGAACCC	AGGACACAAC	6540
CCCAACATGA	CAAACATGAA	TCCACAGTAC	CGAGAAATGG	TGAGGAGACA	GCTGCTACAG	6600
CACCAGCAGC	AGCAGCAGCA	ACAGCAGCAG	CAGCAGCAGC	AACAACAAAA	TAGTGCCAGC	6660
TTGGCCGGGG	GCATGGCGGG	ACACAGCCAG	TTCCAGCAGC	CACAAGGACC	TGGAGGTTAT	6720

GCCCCAGCCA	TGCAGCAGCA	ACGCATGCAA	CAGCACCTCC	CCATCCAGGG	CAGCTCCATG	6780
GGCCAGATGG	CTGCTCCAAT	GGGACAAC'TT	GGCCAGATGG	GGCAGCCTGG	GCTAGGGGCA	6840
GACAGCACCC	CTAATATCCA	GCAGGCCCTG	CAGCAACGGA	TTCTGCAGCA	GCAGCAGATG	6900
AAGCAACAAA	TTGGGTCCACC	AGGCCAGCCG	AACCCCATGA	GCCCCAGCA	GCACATGCTC	6960
TCAGGACAGC	CACAGGCCTC	ACATCTCCCT	GGCCAGCAGA	TCGCCACATC	CCTTAGTAAC	7020
CAGGTGCGAT	CTCCAGCCCC	TGTGCAGTCT	CCACGGCCCC	AATCCCAACC	TCCACATTCC	7080
AGCCCGTCAC	CACGGATACA	ACCCAGCCT	TCACCACACC	ATGTTTCACC	CCAGACTGGA	7140
ACCCCTCACC	CTGGACTCGC	AGTCACCATG	GCCAGCTCCA	TGGATCAGGG	ACACCTGGGG	7200
AACCCCTGAAC	AGAGTGCAAT	GCTCCCCAG	CTGAATACCC	CCAACAGGAG	CGCACTGTCC	7260
AGTGAAGTGT	CCCTGGTTGG	TGATACCACG	GGAGACACAC	TAGAAAAGTT	TGTGGAGGGT	7320
TTGTAG						7326

## (2) INFORMATION FOR SEQ ID NO:14:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2499 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

TCACTTGTC	ATTAATCCAG	CTTCCTTAAT	TTTACTGAAG	AAGAATTTCT	CCAGGATATT	60
GGCACATTTG	TAGTATTCAC	TCTCAGGGGC	GTTGTACTCT	TTGCAATTGG	TAAAGACTCG	120
CTGTAAGTCT	GCCATGAATA	ATTTCTTAGA	CACGTAGTAC	CTATTCTTGA	GGCGTTCACT	180
CATGGTTTTG	AGATCCATGG	GGGACCTTAT	AACCTTCATA	TATCCTGGAG	CTTCTGTCT	240
CTTCACAGGT	TCCATGAAGG	GCCAAGCGCT	TTGATGGCTC	TTCACCTGCT	GGAGGATGCT	300
CTTGAGCGTG	CTGTAAAGCT	GGTCAGGGTC	TCTGGGCTCT	TTACTTTTCT	CTTTTCCACT	360
CGGTTTCCAG	CCTGTCTCTC	TAATTCCAGG	AATGCTTTCT	ATAGGAATCT	GTCGAACTCC	420
ATCTTTAAAA	CATGAAAGTC	CAGGGTAAAC	TTTTCGAATT	TGTGCCTGTT	TTCTTTCAAT	480
CAGTTTTTTA	ATTATCTCCT	TCTGCTTTTT	AATGATGACA	GAAAATTCTG	TGTACGGGAT	540
CCGTGGATTT	AGCTCACATC	CCATTAAAGT	GGCTCCTTCA	TAATCCTTGA	TATAGCCAAC	600
ATATTTGGTT	TTAGGTATTT	TAATTTCTTT	GGAGAAACCC	TGTTTCTTAA	AGTATCCAAT	660
TGCATATTCA	TCTGCATATG	TGAGGAAGTT	CAGGATGTCA	TGCTTTATGT	GATATCTTTT	720
CAAATGATTC	ATCAGGTGTG	TTCCATAGCC	CTTGACTTGC	TCATTTGAGG	TTACAGCACA	780
GAAGACAATC	TCTGTGAATC	CTTGAGATGG	GAACATACGG	AAACAGATAC	CACCAATAAC	840
ACGGCCATCT	TTAATTAAAG	CAAGGGTTTT	GTGTTTCGGG	TCAAAGACGA	GCCGTGTGAT	900
GTATTCTTTT	GGCATTTCGG	GCAGCTGGTG	GGAGAAAACG	TTCTGTAGGC	CAACCAGCCA	960
CATCAGGATC	TTCTTGTTTG	GTTTCTGGTT	GAGGGAATTG	CCAACCACGT	GAAATTCAAT	1020
TACACCCCTG	CGCTCTTCCA	ACCTTGCCGC	CTCATCCCTG	GCCGAGTGTG	CTGACAGAAA	1080
ATTGGTCTCT	GGTCCAAGCA	TTGCTGCAGG	GTCCGTGATG	GTAGACATAA	CCTCGTTGAT	1140
TAATTCATC	GGAATATCCC	CCATAACTCG	GGGTTTCTTG	GCCTCCTCCA	GAACATGAGA	1200
ATCAGTCATT	TTCCCTTTTT	CTCCTGGGTT	TGCCTCAAGT	CCAGAAGAGG	CTTTCAGGC	1260
AGGACTGCTG	CTCCCTGCGT	TTGGCTGCTC	AAGGGAAGAT	GAGGTGAAAT	TGTATGAAAT	1320
TGTCCCAGCC	ACAGGAGGTG	GATTGATAAC	TGTTTGATG	CCTAGCTGGC	TGGTCTGGA	1380
AGAGGCTGAG	AGAAAATCCT	GATCCCAGAT	GGGAGAGTTT	TGACTATATA	CTTCTTCTTC	1440
TAGCATGGAC	AGAAATTTTG	GGAAATGAGT	GAGGATTAGA	GTTCGTTTTT	CAAGAGGCAG	1500
TTTATCTTTT	TCCTGTCTTG	CTTGTTCCAG	GAGTTGTCGC	CTCATAACAG	TGAAGACCGA	1560
GCGAAGCAAT	GTTCTCCCAA	ACACCTGTGT	GGTTTCGTAC	CGAGGTAGAC	TGTCGCAGAA	1620
CTGTGGCACG	TTGCAGTAAC	ACAGCCACCT	TGTGTAGTTC	TCTTTGTATC	CAGAAATATC	1680
ATCATTGGGA	GATCGCAGTC	TTCGTTGAGA	TGGTGCCTCC	AGATGCCAAT	AGTTGATGCG	1740
GTTTAGGAAC	ATTTTGTCCA	ACTCAACTAT	TGTTTGCTTT	TCTTTTGCTG	GCAGGTGACT	1800
AAATTTGTAC	TGCACAAAGT	TATTCACACC	CTGTTCAATG	CTAGGTTTTT	CAAATGGGGG	1860
TTTCTTTTCC	AAAGAGCCTT	CAACCACAGG	TTTTCTCTCT	TGTAATAATAG	ACTTTCTCAA	1920
GAGCTTAAAT	AGATAGAAAT	AAACTTGTTT	GGTATCTGCA	TCTTCTTCCT	TGTGGACACA	1980
GGTAAAGAGA	TATTCACAT	CCAATCAAT	TCCCAGGAGT	CTGTTCAATT	CTTCTCTGTA	2040
CACATTCTCC	AGGTGGGAAA	CATGAGCAGC	TAGGGCATGG	CTACAACTCC	GACAGGATTC	2100
TGTTAGACTG	ACAATTATTT	GCTGCAGGTC	GGCTCTGGGG	GGAGTGGGTG	AGGGGTTAGG	2160
GTTTTTCCAG	CCATTACATT	TACAAGACTC	CTCGGCCTTG	CAGGCGGAGT	ACACTCCGAG	2220
TTTCTCCAGT	TTCTTGCCCC	GCGGAGCGGA	GCGTAGTTGC	GCTTTCTTCA	CGGCGATTCG	2280
GGCCGAGCCA	CCGCCTCCCG	GTCCTTCGGC	CGTGCCCGCT	GCAGCCACTG	CCGTCGCCGG	2340
ACCGCAGGCG	CCCGAGCCCC	CGGCGGCAGC	GCGCAGGGG	GAGCCCTGCG	GGGGCGCGGG	2400

CGGAAGCGCC GCAGGCTGCG GGGGCAGCGC CCCGGGCCCCG GCCCCTGCCC CGGCTCCTGC 2460  
 CCCGCAGCCG CCCGGCCCCG CCCC GCCAGC CTCGGACAT 2499

## (2) INFORMATION FOR SEQ ID NO:15:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2442 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

TC	ACTTGTCA	ATCA	ACCCTG	CTTC	CTTAAT	TTT	ACTGAAG	AAGA	ACTTCT	CC	AGGATGCT	60	
GG	CGCATT	TTG	TAGT	ACTCGC	TCT	CGGAGG	GTT	GTA	CTTGG	TG	AACATCG	120	
TT	GCAAGT	TCC	GCCAT	GAATA	ACT	TCTTAGA	CAC	ATAGT	AC	CTG	TTCCTGA	180	
CAT	GGTTTT	AG	ATCCAT	GG	GGA	ACCCTTAT	AA	CTTCATAA	TAT	CCCCGGAG	CTT	CTGTTCT	240
CTT	CACTGGT	TCC	ATGAAAG	GCCA	AAGCATT	TGG	ATGGTTC	TT	CACCTGCT	GC	AGGATGTT	300	
CTT	GAGGGTG	CTG	TAAACGT	GCT	CAGGGTC	TTT	GGGCTCT	TT	ACTTTTTCT	CT	TTTCCACT	360	
TGG	TTTCCAG	CCT	GCTCTC	TG	ATTCAGG	AAT	GCTTTCT	AT	AGGAATCT	GCC	GAACTCC	420	
AT	CTTTGAAA	CAC	GAAAGTC	CAG	GGTAGAC	TTTT	CGAATC	TGG	GCTTGT	TT	CTTTCTAT	480	
CAG	CTTTTTA	ATG	ATCTCCT	TCT	GTCTTTT	AAT	GATGACA	GAG	AACTCTG	TG	TATGGGAT	540	
CTG	AGGGTTC	AG	CTCACAT	CC	ATCAAAGT	GG	CCCCTTCA	TA	ATCCTTGA	TG	TAGCCAAC	600	
AT	ATTTGGTT	TT	AGGTATTT	TG	ATTTCTTT	GG	AGAAACCC	TG	CTTCTTGA	AA	TAGCCGAT	660	
GG	CATACTCA	TCT	GCAATG	TG	AGGAAGTT	GAG	GATCTCG	TG	CTTTATGT	GG	TATCTTT	720	
GAG	ATGGTTC	ATC	AGGTGGG	TT	CCATAGCC	CTT	GACTTGT	TC	ATTTGAGG	TT	ACTGCACA	780	
GAA	ACAATC	TCT	GTGAATC	CCT	GGGATGG	AA	CATCCGG	AA	CAGATAC	CAC	CAATGAC	840	
AC	GGCCATCT	TTA	ATTAAAG	CA	AGGTTTTT	GT	GTTTCGGG	TC	AAAGACGA	GCC	GTGTGAT	900	
GT	ACTCTTTG	GG	CATTCTGG	GC	AGCTGGTG	GG	AAACACA	TT	CTGGAGGC	CC	ACGAGCCA	960	
CAT	CAGGATC	TT	CTTGTTT	GT	TTCTGGTT	CAG	GGAGTTG	CCC	ACCACGT	GG	AATTCAAT	1020	
GAC	ACCCCTG	CG	TTCTTCCA	GCC	GTGCCGC	CT	CATCTCTG	GCC	GATGGG	CT	GACAGAAA	1080	
ATT	GGTCTCT	GG	TCCAAGCA	TCC	CTGCAGG	GT	CTGTGATG	GT	AGACATGA	CCT	CATTGAT	1140	
CA	ATTCCACG	GG	AATATCCC	CC	ATCACTCG	AG	ATCTCTTG	GC	CTCCTCGG	GAG	CATGAGA	1200	
GTT	GTTCATT	TT	CCTCTTTT	CT	CCGGGTT	TG	CTTCAAGC	CC	AGAAGAGC	CT	CTGCATCC	1260	
AG	GACTTGTT	CT	CCCTCCAT	TG	ATCTGCTC	AT	GGGAAGTT	GA	ATTTGAAC	TG	AACAATGC	1320	
TGT	CCCAGTA	AC	AGGAGGAC	TG	ATTACTGT	TT	GGATTCT	AG	CGGGCTGG	TT	CTGGAAGA	1380	
GG	CTGAGAGA	AA	ATCCTGAT	CCC	CAGATAGG	AG	AATTTTGA	CT	ATACACTT	CT	TCTTCCAA	1440	
CAT	GGACAGA	AA	CTTTGGGA	AA	TGTGTGAG	GAT	AAGCGTG	CG	TTTCTCAA	GAG	GCAGTTT	1500	
GT	CTTTTTT	TG	TCTGGCTT	GT	TCCAAGAG	CT	GTCTCTC	AT	GATGGTGA	AG	ACCGAGCG	1560	
AAG	CAATGTT	CT	CCCAAACA	CCT	TTGTGGT	TT	CTTACCGA	GG	TAACTGT	CAC	AGAAGT	1620	
CG	GATACATT	G	CAGTAGCACA	ACC	ACCTTGT	GT	AGTTTTTCC	TT	GATCCAG	AG	ATGTCATC	1680	
ATT	GGGAGAC	CG	TAGTCTCC	GCT	GAGATGG	AG	CCTCCAGA	TG	CCAGTAGT	TG	ATGCGGTT	1740	
CAG	AAACATC	TT	GGCCAGCT	CG	ATCGTTGT	CT	GCCTCTCT	TT	CGATGGCA	AG	TGACTAAA	1800	
CTT	GTA	CTG	C	ACGAAGTTGT	TC	ACACCCTG	TT	CAATACTG	GG	CTTCTCAA	AT	GGCGGCTT	1860
CTT	TCTCAAG	GAG	CCTTCAA	CC	ACAGGTTT	TC	CTCTTTGT	AAA	ATTGACT	TT	CTCAAGAG	1920	
CTT	GAA	TAG	AGGTACA	CTT	GTTTGGT	AT	CTGCATCT	TCT	TCTTTGT	GG	ACGCAGGT	1980	
GA	AGGTTAC	TCC	ACATCCA	AC	ACAATTCC	CAG	GAGTCTG	TCC	ATCTCTT	CCT	CTGACAC	2040	
ATT	TCTCAAG	TG	AGAAACGT	GAG	CAGCAAG	GG	CATGGCTA	CAG	CTTCGAC	AG	GATTCTGT	2100	
CAA	ACTGACA	ATT	ATCTGCT	GG	AGGTCTCC	TCT	TGGTGGA	GT	AGGAGAGG	GG	TAGGGTT	2160	
CTT	CCAGCCA	TT	GATTTTAC	AG	GA	CTCTC	TG	CGTTGCAG	GC	GGAGTACA	CG	CCGAGTTT	2220
CT	CCAGCTTC	TT	CGCCCGCG	GAG	CAGAGCG	CA	ACTGCGCC	TT	CTTACGCG	CG	ATCCGGGC	2280	
CG	AGCCGCCT	CCT	CCCGGTC	CCT	CGCGGT	GCC	CGCCGCG	GCC	ACCGGCG	TC	GCTGGCCC	2340	
GC	AGGAAGCA	GAG	CTCCCGG	CAG	CGGTGGC	CAG	GGTCCGG	GG	GGAACCGT	GC	GGGGGCGC	2400	
GG	GAGGCAGT	GCT	GGGGACC	CG	GCCCCGCC	AG	CCTCGGCC	AT				2442	

## (2) INFORMATION FOR SEQ ID NO:16:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CCCCGCCAGCC TCGGACATGC

20

## (2) INFORMATION FOR SEQ ID NO:17:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

CCCCGCCAGCC TCGGCCATGC

20

## (2) INFORMATION FOR SEQ ID NO:18:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2442 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

ATGGCCGAGG	CTGGCGGGGC	CGGGTCCCCA	GCACTGCCTC	CCGCGCCCCC	GCACGGTTCC	60
CCCCGGACCC	TGGCCACCGC	TGCGGGGAGC	TCTGCTTCCT	GCGGGCCAGC	GACGCCGGTG	120
GCCGCGGCGG	GCACCGCCGA	GGGACCGGGA	GGAGGCGGGT	CGGCCCCGAT	CGCCGTGAAG	180
AAGGCGCAGT	TGCGCTCTGC	TCCGCGGGCG	AAGAAGCTGG	AGAAACTCGG	CGTGACTTCC	240
GCCTGCAAGG	CAGAGGAGTC	CTGTAAATGC	AATGGCTGGA	AGAACCCTAA	CCCCCTCTCT	300
ACTCCACCAA	GAGGAGACCT	CCAGCAGATA	ATTGTCAATT	TGACAGAATC	CTGTCGAAGC	360
TGTAGCCATG	CCCTTGCTGC	TCACGTTTCT	CACTTGGAGA	ATGTGTCAGA	GGAAGAGATG	420
GACAGACACC	TGGGAATTGT	GTTGGATGTG	GAGTACCTCT	TCACCTGCGT	CCACAAAGAA	480
GAAGATGCAG	ATACCAAACA	AGTGTACTTC	TACCTATTCA	AGCTCTTGAG	AAAGTCAATT	540
TTACAAAGAG	GAACACCTGT	GTTTGAAGGC	TCCTTGGAGA	AGAAGCCGCC	ATTTGAGAAG	600
CCCAGTATTG	AACAGGGTGT	GAACAACCTC	GTGCAGTACA	AGTTTAGTCA	CTTGCCATCG	660
AAAGAGAGGC	AGACAACGAT	CGAGCTGGCC	AAGATGTTTC	TGAACCGCAT	CAACTACTGG	720
CATCTGGAGG	CTCCATCTCA	GCGGAGACTA	CGGTCTCCCA	ATGATGACAT	CTCTGGATAC	780
AAGGAAAACT	ACACAAGGTG	GTTGTGCTAC	TGCAATGTAC	CGCAGTTCTG	TGACAGCTTA	840
CCTCGGTACG	AAACCACAAA	GGTGTTTGGG	AGAACATTGC	TTCGCTCGGT	CTTCACCATC	900
ATGAGACGAC	AGCTCTTGGA	ACAAGCCAGA	CAGAAAAAAG	ACAAACTGCC	TCTTGAGAAA	960
CGCACGCTTA	TCCTCACACA	TTTCCCAAAG	TTTCTGTCCA	TGTTGGAAGA	AGAAGTGAT	1020
AGTCAAAAT	CTCCTATCTG	GGATCAGGAT	TTTCTCTCAG	CCTCTTCCAG	AACCAGCCCCG	1080
CTAGGAATCC	AAACAGTAAT	CAGTCCTCCT	GTTACTGGGA	CAGCATTGTT	CAGTTCAAAT	1140
TCAACTTCCC	ATGAGCAGAT	CAATGGAGGG	AGAACAAGTC	CTGGATGCAG	AGGCTCTTCT	1200
GGGCTTGAAG	CAAACCCGGG	AGAAAAGAGG	AAAATGAACA	ACTCTCATGC	TCCCAGAGGAG	1260
GCCAAGAGAT	CTCAGTGTAT	GGGGGATATT	CCCGTGGAA	TGATCAATGA	GGTCATGTCT	1320
ACCATCACAG	ACCCCTGCAGG	GATGCTTGGA	CCAGAGACCA	ATTTTCTGTC	AGCCCATTCG	1380
GCCAGAGATG	AGGCGGCACG	GCTGGAAGAA	CGCAGGGGTG	TCATTGAATT	CCACGTGGTG	1440
GGCAACTCCC	TGAACCAGAA	ACCAAACAAG	AAGATCCTGA	TGTGGCTCGT	GGGCCTCCAG	1500
AATGTGTTTT	CCCACCAGCT	GCCCAGAATG	CCCAAAGAGT	ACATCACACG	GCTCGTCTTT	1560
GACCCGAAAC	ACAAAACCCT	TGCTTTAATT	AAAGATGGCC	GTGTCAATTG	TGGTATCTGT	1620
TTCCGGATGT	TTCCATCCCA	GGGATTACAA	GAGATTGTTT	TCTGTGCAGT	AACCTCAAAT	1680
GAACAAGTCA	AGGGCTATGG	AACCCACCTG	ATGAACCATC	TCAAAGAATA	CCACATAAAG	1740
CACGAGATCC	TCAACTTCCT	CACATATGCA	GATGAGTATG	CCATCGGCTA	TTTCAAGAAG	1800
CAGGGTTTCT	CCAAAGAAAT	CAAAATACCT	AAAACCAAAT	ATGTTGGCTA	CATCAAGGAT	1860
TATGAAGGGG	CCACTTTGAT	GGGATGTGAG	CTGAACCCTC	AGATCCCATA	CACAGAGTTC	1920
TCTGTCATCA	TTAAAAAGCA	GAAGGAGATC	ATTAAAAAGC	TGATAGAAAG	AAAACAAGCC	1980
CAGATTCGAA	AAGTCTACCC	TGGACTTTCC	TGTTTCAAAG	ATGGAGTTCC	GCAGATTCTC	2040

ATAGAAAGCA	TTCCTGGAAT	CAGAGAGACA	GGCTGGAAAC	CAAGTGGAAA	AGAGAAAAGT	2100
AAAGAGCCCA	AAGACCCTGA	GCACGTTTAC	AGCACCCCTCA	AGAACATCCT	GCAGCAGGTG	2160
AAGAACCATC	CAAATGCTTG	GCCTTTCATG	GAACCAGTGA	AGAGAACAGA	AGCTCCGGGA	2220
TATTATGAAG	TTATAAGGTT	CCCCATGGAT	CTGAAAACCA	TGAGTGAACG	CCTCAGGAAC	2280
AGGTACTATG	TGTCTAAGAA	GTTATTCATG	GCGGACTTGC	AACGAGTGTT	CACCAACTGC	2340
AAGGAGTACA	ACCCTCCCGA	GAGCGAGTAC	TACAAATGCG	CCAGCATCCT	GGAGAAGTTC	2400
TTCTTCAGTA	AAATTAAGGA	AGCAGGGTTG	ATTGACAAGT	GA		2442

**What is claimed is:**

1. A purified protein designated P/CAF having a molecular weight of about 93,000 daltons as determined by sodium dodecyl sulfate polyacrylamide gel electrophoresis under reducing conditions and which acetylates histones.
2. The protein of claim 1 consisting of the amino acid sequence of SEQ ID NO:1.
3. The protein of claim 1 comprising the amino acid sequence of SEQ ID NO:2.
4. The protein of claim 1, which also binds to the amino acid sequence of SEQ ID NO:3 on a p300 cellular protein and to amino acid residues 1805-1854 of a CBP cellular protein (SEQ ID NO:9).
5. A fragment of the protein of claim 1 having histone acetyltransferase activity.
6. A polypeptide consisting of the amino acid sequence of SEQ ID NO: 2.
7. A fragment of the protein of claim 1 which binds to the amino acid sequence of SEQ ID NO: 3 on the p300 cellular protein and the amino acid sequence of SEQ ID NO:9 on the CBP cellular protein.
8. A polypeptide consisting of the amino acid sequence of SEQ ID NO:4.
9. A nucleic acid consisting of the nucleotide sequence of SEQ ID NO:10.
10. A nucleic acid having a nucleotide sequence which encodes the protein of claim 1.

11. A nucleic acid having a nucleotide sequence which encodes the protein of claim 2.
12. A nucleic acid having a nucleotide sequence which encodes the protein of claim 3.
13. A nucleic acid consisting of the nucleotide sequence which encodes the protein of claim 4.
14. A nucleic acid complementary to and which selectively hybridizes with the nucleic acid of claim 11 under stringent hybridization conditions.
15. A fragment of the nucleic acid of claim 9, which encodes a polypeptide that acetylates histones.
16. A fragment of the nucleic acid of claim 9, which encodes a polypeptide which binds to the amino acid sequence of SEQ ID NO:3 on the p300 cellular protein and the amino acid sequence of SEQ ID NO:9 on the CBP cellular protein.
17. A purified antibody which specifically binds the protein of claim 1.
18. A purified antibody which specifically binds the protein of claim 2.
19. A purified antibody which specifically binds the protein of claim 3.
20. A purified antibody which specifically binds the protein of claim 4.
21. An assay for screening substances for the ability to inhibit or stimulate the histone acetyltransferase activity of P/CAF comprising:
  - a) contacting the substance with a system in which histone acetylation by P/CAF can be determined;

b) determining the amount of histone acetylation by P/CAF in the presence of the substance; and

c) comparing the amount of histone acetylation by P/CAF in the presence of the substance with the amount of histone acetylation by P/CAF in the absence of the substance, a decreased or increased amount of histone acetylation by P/CAF in the presence of the substance indicating a substance that can inhibit or stimulate, respectively, the histone acetyltransferase activity of P/CAF.

22. An assay for screening substances for the ability to inhibit binding of P/CAF to p300/CBP comprising:

a) contacting the substance with a system in which the P/CAF binding of P300/CBP can be determined;

b) determining the amount of P/CAF binding of p300/CBP in the presence of the substance; and

c) comparing the amount of binding of P/CAF to p300/CBP in the presence of the substance with the amount of binding of P/CAF to p300/CBP in the absence of the substance, a decreased amount of binding of P/CAF to p300/CBP in the presence of the substance indicating a substance that can inhibit the ability to inhibit binding of P/CAF to p300/CBP.

23. The method of claim 22, wherein the system consists of a cell free reaction mixture comprising a fragment of the p300 protein comprising amino acid residues 1767-1816 (SEQ ID NO:3) and the protein of claim 4.

24. The method of claim 22, wherein the system consists of a cell free reaction mixture comprising a fragment of the CBP protein comprising amino acid residues 1805-1854 (SEQ ID NO:9) and the protein of claim 4.

25. The method of claim 22, wherein the system consists of a cell extract produced from cells producing both p300 and P/CAF.

26. An assay for screening substances for the ability to inhibit or stimulate the histone acetyltransferase activity of p300/CBP comprising:

- a) contacting the substance with a system in which histone acetylation by p300/CBP can be determined;
- b) determining the amount of histone acetylation by p300/CBP in the presence of the substance; and
- c) comparing the amount of histone acetylation by p300/CBP in the presence of the substance with the amount of histone acetylation by p300/CBP in the absence of the substance, a decreased or increased amount of histone acetylation by p300/CBP in the presence of the substance indicating a substance that can inhibit or stimulate, respectively, the histone acetyltransferase activity of p300/CBP.

27. An assay for screening substances for the ability to inhibit binding of a DNA-binding transcription factor to p300/CBP comprising:

- a) contacting the substance with a system in which the DNA-binding transcription factor binding of p300/CBP can be determined;
- b) determining the amount of DNA-binding transcription factor binding of p300/CBP in the presence of the substance; and
- c) comparing the amount of binding of DNA-binding transcription factor to p300/CBP in the presence of the substance with the amount of binding of DNA-binding transcription factor to p300/CBP in the absence of the substance, a decreased amount of binding of DNA-binding transcription factor to p300/CBP in the presence of the substance indicating a substance that can inhibit the ability to inhibit binding of DNA-binding transcription factor to p300/CBP.

28. The method of claim 27, wherein the system consists of a cell free reaction mixture comprising a DNA-binding transcription factor and p300/CBP.

29. The method of claim 27, wherein the system consists of a cell free reaction mixture comprising a fragment of the CBP protein comprising a DNA-binding transcription factor and p300/CBP.

30. The method of claim 27, wherein the system consists of a cell extract produced from cells producing both a DNA-binding transcription factor and p300/CBP.

31. The method of claim 27, wherein the DNA-binding transcription factor is selected from the group consisting of a nuclear hormone receptor, CREB, c-Jun/v-Jun, c-Myb/v-Myb, YY1, Sap-1a, c-Fos, MyoD and SRC-1.

32. A method for inhibiting the transcription modulating activity of P/CAF in a subject, comprising administering to the subject a transcription modulating activity inhibiting amount of a substance in a pharmaceutically acceptable carrier.

33. The method of claim 32, wherein the substance can inhibit the transcription modulating activity of P/CAF by preventing the binding of P/CAF to p300/CBP.

34. A method for stimulating the transcription modulating activity of P/CAF in a subject, comprising administering to the subject a transcription modulating activity stimulating amount of a substance in a pharmaceutically acceptable carrier.

35. The method of claim 34, wherein the substance can stimulate the transcription modulating activity of P/CAF by promoting the binding of P/CAF to p300/CBP.

36. The method of claim 34, wherein the substance can stimulate the transcription modulating activity of P/CAF by stimulating the histone acetyltransferase activity of P/CAF.

37. A method for inhibiting the histone acetyltransferase activity of p300/CBP in a subject, comprising administering to the subject a histone acetyltransferase activity inhibiting amount of a substance in a pharmaceutically acceptable carrier.

38. The method of claim 37, wherein the substance can inhibit the transcription modulating activity of p300/CBP by preventing the binding of a DNA-binding transcription factor to p300/CBP.

39. The method of claim 38, wherein the DNA-binding transcription factor is selected from the group consisting of a nuclear hormone receptor, CREB, c-Jun/v-Jun, c-Myb/v-Myb, YY1, Sap-1a, c-Fos, MyoD and SRC-1.

40. The method of claim 37, wherein the substance is an antibody which binds p300/CBP.

41. A method for stimulating the histone acetyltransferase activity of p300/CBP in a subject, comprising administering to the subject a histone acetyltransferase activity stimulating amount of a substance in a pharmaceutically acceptable carrier.

42. The method of claim 41, wherein the substance can stimulate the histone acetyltransferase activity of p300/CBP by promoting the binding of a DNA-binding transcription factor to p300/CBP.

1/4

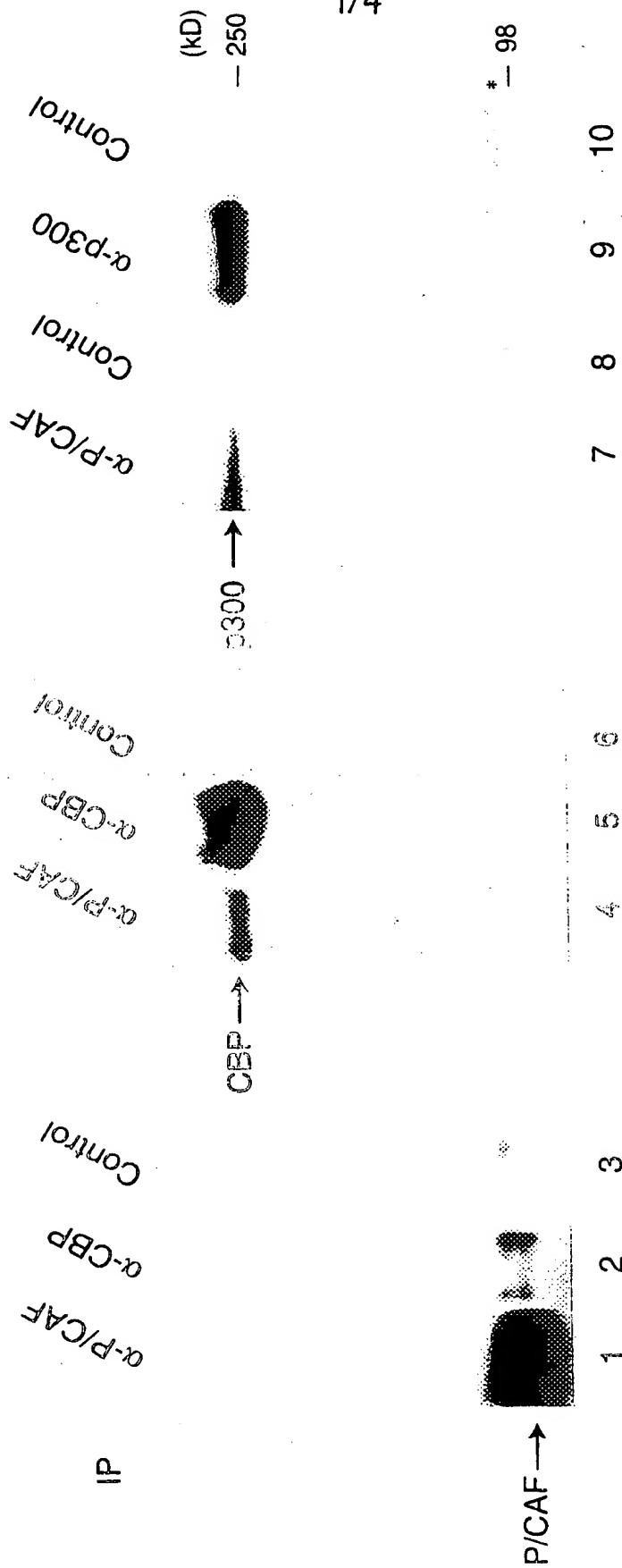


FIG.1A

FIG.1B

FIG.1C

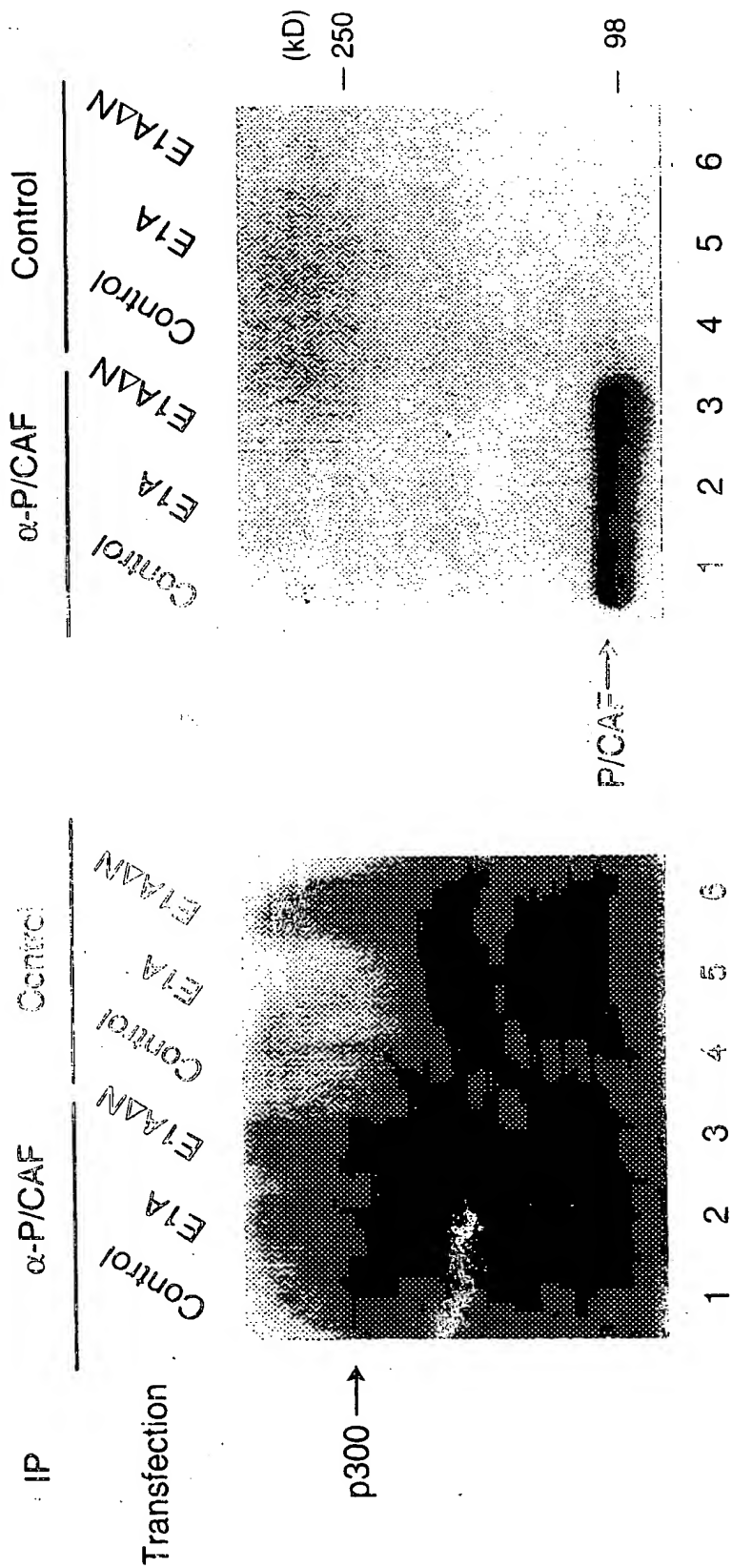


FIG.1E

FIG.1D

3/4

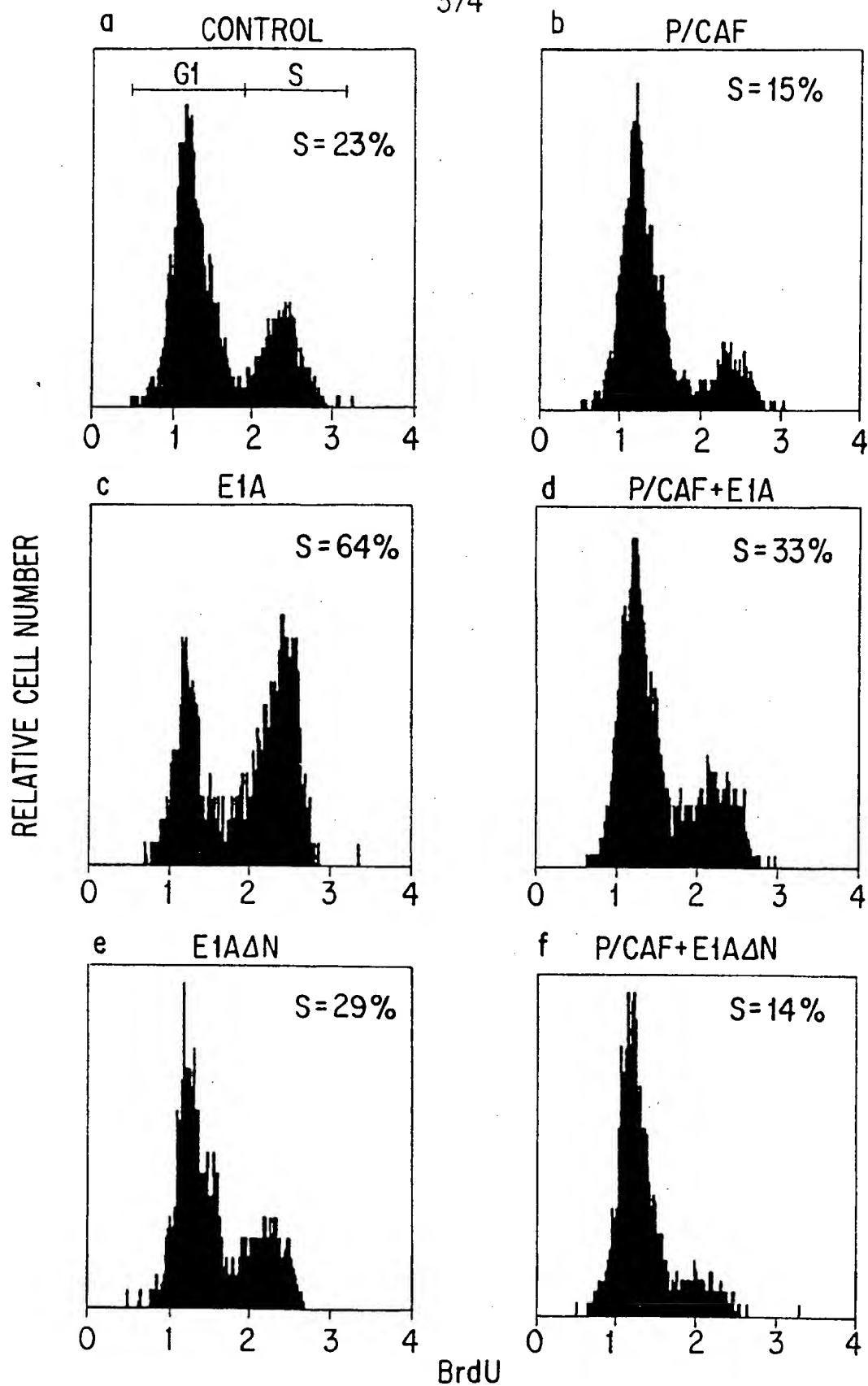


FIG. 2

4/4

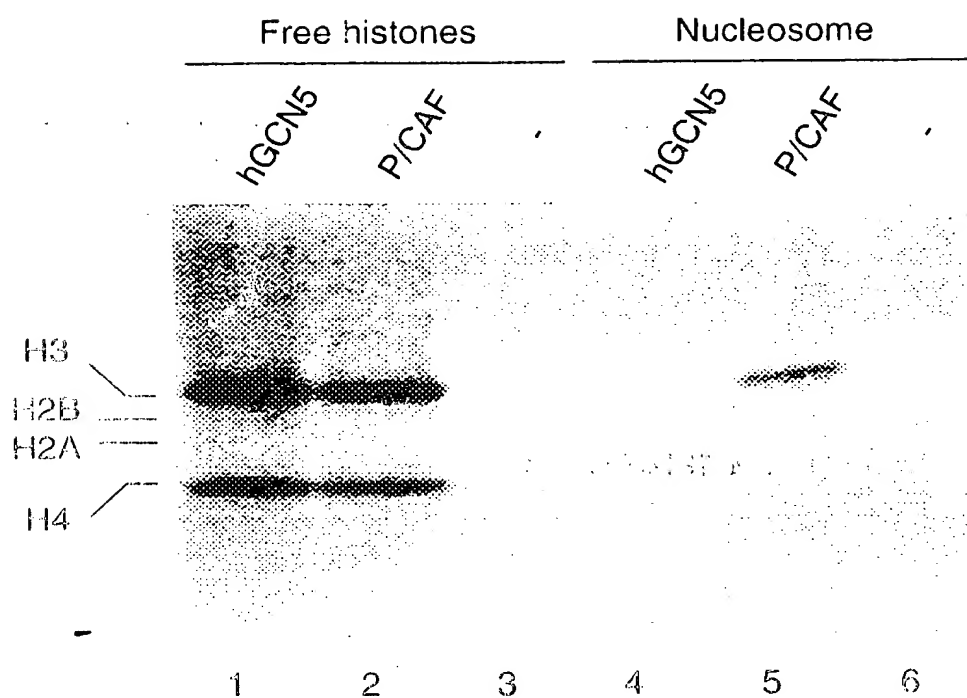


FIG. 3

# INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 97/12877

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/12 C07K14/47 G01N33/50 A61K38/17

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EMBL EST, Accession number N39522 Sequence no yv27b08.s1 Homo sapiens cDNA clone 243927 3' 25 January 1996 XP002050402 see the whole document	1
A	GEORGAKOPOULOS, T. & THIREOS, G.: EMBO JOURNAL., vol. 11, 1992, EYNSHAM, OXFORD GB, pages 4145-4152, XP002050399 see the whole document	1



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

### \* Special categories of cited documents :

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

\*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

\*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

\*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

\*Z\* document member of the same patent family

Date of the actual completion of the international search

17 December 1997

Date of mailing of the international search report

14. 01. 98

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Chambonnet, F

# INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 97/12877

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.:
P,X	<p>YANG, X.Y. ET AL.: "A p300-CBP-associated factor that competes with the adenoviral oncoprotein E1A"</p> <p>NATURE., vol. 382, no. 8589, 25 July 1996, LONDON GB, pages 319-324, XP002050400 see the whole document</p>	1
P,X	<p>OGRYZKO, V.V. ET AL.: "The transcriptional coactivators p300 and CBP are histone acetyltransferases"</p> <p>CELL, vol. 87, no. 5, November 1996, NA US, pages 953-959, XP002050401 see the whole document</p>	1
P,X	<p>EMBL EST, Accession number U57316, Sequence reference human GCN5 (hGCN5) complete cds. 26 august 1996 XP002050403 see the whole document</p>	1

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US 97/12877

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:  
see FURTHER INFORMATION sheet PCT/ISA/210
2. ☐ Claims Nos.:  
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

### Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Remark : Although claims 32 to 42 are directed to a method of treatment of the human/animal body , the search has been carried out and based on the alleged effects of the composition.